A Scenario Generation Algorithm for Multistage Stochastic Programming: Application for Asset Allocation Models with Derivatives

Alessandro Laurent

Submitted for the degree of Ph.D. in Economics at

Faculty of Economics, Department of Finance
University of Lugano
Lugano, Switzerland

Thesis Committee:
Prof. G. Barone-Adesi, advisor, University of Lugano
Prof. W. Ziemba, University of British Columbia
Prof. F. Audrino, University of Lugano

March 2006

To my family

"What makes the future different from the past is the choice that the participants are obliged (and privileged) to exercise on the basis of their imperfect understanding."

George Soros, Open Society.

Acknowledgments

This work was influenced by many people I had the pleasure to meet in the last years. My gratitude goes to everyone who directly or indirectly helped me during the completion of this PhD thesis.

First of all I would like to express my gratitude to Prof. Dr. Giovanni Barone Adesi for his willingness to advise this thesis, and for the patience demonstrated in these years. This work would not have been possible without his advice and help. My gratitude also goes to Prof. Dr. Francesco Audrino for his helpfull and precise comments and suggestions and for accepting to be a member of the thesis committee. Moreover, I really wish to thank Prof. Dr. William T. Ziemba for accepting to be the external member of my thesis committee.

For the stimulating discussions on the most diverse subjects and for his helpful and motivating comments on this thesis, a huge thanks to Fulvio especially, but also to my PhD fellow travellers Valeria, Ombretta, Jens, Jost and Luigi.

For their friedship and support during the last phase of this work my gratitude also goes to my collegues and friends in Thalia, the fund of hedge funds in which I moved my first stimulating and pleasant steps in the working world.

For the careful English proof-reading, many thanks to Claudia and John.

Last but not least, I would like to express my most sincere gratitude to my family for their continuous encouragement, support and constant presence. A special tanks goes to my brother Andrea, who suggested me the thesis argument and was a helpful and indispensable guide during these last years.

Financial support from Fondazione Ceresio is gratefully acknowledged.

Lugano, March 2006 Alessandro Laurent

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation and objective

"The process of selecting a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio". Harry Markowitz opened his famous 1952 article about portfolio selection with this sentence. In over fifty years financial markets have drastically evolved and a huge amount of new financial instruments have arisen. Nevertheless Markowitz's statement is still correct, as it has been for fifty years. Today, like fifty years ago, in order to select a portfolio optimally, we have to construct the future performances of the available securities. Once we have established these performances, defined the objective function we want to maximize and the conditions under which the maximization occurs, we may select the portfolio.

The Markowitz theory of portfolio selection assumes that returns are normally distributed and investors have risk-averse utility functions. Therefore an asset's value depends on the expectations, the variance of its return and the covariances of its return with the returns of all other existing and potential investments. The assumption of normality allows the variance (and the related covariances) to be the only source of risk. Moreover it is a static, single period model based on a buy-and-hold strategy. Thus, given an investment opportunity set, we are able to find the set of portfolios which maximize expected returns for any desired level of variance. This set is called the *efficient frontier*.

Is the Markowitz framework today still valid after fifty years? Given the evolution of financial markets, can we be satisfied with the classical mean-variance setup? The answer seems to be no. The main feature of investment and financial problems is the necessity to take decisions *under uncertainty* and *over more than one period* (Dupačová (1999)). These two aspects represent the main limitations of the Markowitz mean-variance model. First, using the variance as the unique source of uncertainty is too restrictive. Besides the risk of the portfolio other possible factors of uncertainty like liquidity constraints, transaction costs, future interest rates, among others, are important aspects which should be considered in the decision-making process of defining the best structure of a portfolio. Second, the introduction of options in the investment opportunity set leads the Markowitz approach to be rejected. Because of the intrinsic structure of options, which cause the distribution of the return of the underlying asset to be followed only up to a certain 'strike' level, we cannot simply introduce options in the Markowitz mean-variance setup and compute the best portfolio for a given level of risk. Variance is in this case an inappropriate source of uncertainty, and the concept of risk has to be redefined. Finally, to base our decision on a single period maximization problem repeated over more than one period can be far from a good suboptimal dynamic decision (Carino, Myers, and Ziemba (1998)). Portfolio selection is normally subject to long term goals. In order to achieve these goals the portfolio is readjusted over time depending on how the market evolved. In a static single period model, like the Markowitz one, every period we try to make the best choice given the expectations regarding the future. In a dynamic multiperiod model we go one step further. We try to make the best decision considering not only the possible future realisations of the random variables, but also the possible choices which could be

made in each future period.

To solve these problems a new portfolio selection approach has been developed. According to Carino and Ziemba (1998), this should be superior to the static Markowitz-style portfolio selection, which despite being the industry standard for asset allocation, has several limitations for asset-liability management over time. In particular it should have more realistic definitions of operating risk and business constraints than return variance, as well as a more explicit description of business goals, a better definition of uncertainty, and a multistage as opposed to a single period analysis. This new approach is based on multiperiod (multistage) stochastic programming. The Markowitz static single-period maximization of portfolio return for every level of variance is replaced by a dynamic multinomial maximization of the expected net present value (ENPV) under a flexible set of constraints (Hester and Pierce (1975)). As mentioned above, the decision process is dynamic in the sense that the optimal first period decision depends on the actions which will be taken in each future period for each uncertain event (Bradley and Crane (1972)).

Portfolio allocation models using a multiperiod dynamic criterion can be *deterministic* or *stochastic*. Deterministic models use linear programming, and assume particular realisations for random events. Their applications in portfolio management date back to Cohen and Hammer (1967) and were widely used in the 1960s-1970s because of their computational tractability for large problems. Probability distributions can be obtained only for different economic realisations and a linear programming formulation can be applied to each realisation to determine optimal solutions. However, this approach will not generate an optimal solution to the problem as a whole, but rather act as a deterministic simulation to observe portfolio behaviour under various economic conditions (Kusy and Ziemba (1986)). Multiperiod stochastic programming is much more complex since random events do not assume particular realisations, but multidimensional data trajectories in a probability space. Instead of using a particular realisation for each random variable (for instance the expected value), an entire set of possible events (scenarios) and their probabilities is considered. The cost of considering such multidimensional possible realisations of random events in the optimization is due to inherent computational difficulties. This (and the limitations of computer technology) is why stochastic models based on multidimensional scenarios were not very popular until the 1980s. The first application of multistage stochastic programs in portfolio allocation and management dates back to Bradley and Crane (1972), but it is only in the 1980s that an important diffusion of this approach is witnessed, e.g. Brodt (1984), Kusy and Ziemba (1986), Dempster and Ireland (1988), Mulvey and Vladimirou (1989), among others. In recent years, with the progress in numerical methods, software and computer technologies, the applications of stochastic programming have become much more complex and realistic (see for example Carino, Myers, and Ziemba (1998) and Carino and Ziemba (1998)). According to Dupačová (1999) the strength of stochastic programming is in open possibilities to support the asset and liability management and the risk management decisions under various circumstances which reflect the goals and the restrictions of the users.

Due to its flexibility, such a scenario-based stochastic program seems to be the optimal framework to introduce contingent claims in the portfolio selection problem. The introduction of options would seem a natural extension of the existing asset-liability models cited above which are based on stochastic programming and consider only cash, stocks and bonds in their investment opportunity set. To the best of our knowledge, however,

no model (apart from Laurent (2003)) has ever taken on this task. The reason is probably to be found in the difficulty of generating consistent scenarios for option returns. As mentioned above, random events enter in a stochastic program as multidimensional trajectories in a probability space. For every random variable at each period in time a limited number of events and the probabilities of their occurrence have to be computed. Over the entire investment period considered, this set of intertemporally connected events is described by a scenario tree. For every asset considered in the portfolio we have to define the future prices (returns) and the respective probabilities at every node of the tree. The optimal solution depends crucially on the choice of these scenario trees, which should describe the future probability distributions of all considered random variables in the best possible way. While several techniques already exist for the generation of stock and bond scenarios, to our knowledge there are no methodologies which allow to define a flexible number of events, consistent with the respective underlying scenarios, for contingent claims. The challenge consists in finding a flexible methodology for pricing options at every node of the scenario tree, in order to maintain today's option market prices and, at expiration, to keep these prices consistent with underlying and strike prices. Solving this problem would allow us to introduce options in a stochastic program, enlarge the investment opportunity set and provide a better portfolio selection model. This is the objective of our work.

In order to reach this objective, the structure of our model can be synthesized in the following way: first, define the probability distribution function of the underlying assets considered in the portfolio. Second, discretize these distributions at each investment date in order to obtain a limited number of nodes and, over the entire investment period, a scenario tree. Since the scenarios crucially depend on the distribution functions, these have to be defined very carefully. We will not hypothesise a functional form for the distributions, but instead simulate them by a GARCH process. This allows us to consider volatility clustering and the correlation among the different assets. Moreover, the intertemporal dependence between subsequent nodes of the tree can be easily inferred by conditioning the next step distributions to the different nodes obtained on the previous investment date. Finally, once we have the entire scenario tree for the underlying assets, we can compute the scenario tree for the options. Under the hypothesis that the options all expire at the end of the investment period, we can infer the prices for every node of these trees, in such a way that the theoretical prices are as close as possible to the observed option market prices, that the structure of the underlying scenario tree is maintained and that no arbitrage possibility is allowed. Once obtained the scenario trees for all underlying assets and options, can be used as input to a multistage stochastic program in order to build the best strategy for portfolio allocation.

1.2 Structure

This thesis is divided into two main parts. Part I, presents the theoretical aspects needed to capture the problems we are faced with and to understand the different steps of the scenario generation algorithm.

In section 2 we start by describing how a dynamic stochastic program is formulated and solved, and its advantages with respect to a static, deterministic optimization model like Markowitz's. We will explain the importance of properly defining future scenarios

and in particular how the optimal solution of the program is sensitive to the description of possible future random events.

In section 3 we show how a limited number of scenarios may be established given a well known probability density function, which describes the distribution of future random variables. We present different discretization methodologies and try to identify if there exists a superior discretization procedure.

Section 4 deals with the method used to infer future stock price distributions. As previously introduced, instead of imposing a functional form for the price density, we simulate it from past data by a GARCH process. We present three different GARCH processes and explain the characteristic and advantages for each of these processes. Furthermore, the simulation procedure is explained.

To conclude the theoretical part, Section 5 describes the generation of option prices' scenarios. Because of their particular structure, the generation of option price scenarios cannot be computed as for other random variables like underlying asset prices. Many conditions linking the underlying asset and the related option have to be satisfied.

In part II we propose the algorithm for the generation of the scenario trees for both the underlying assets and the related options. Some empirical results are presented. We analyze the DAX 100 index from January 1991 to May 2004 and compute the one month scenario trees starting on April 23, 2004 for the index and for 26 European call and 26 European put options on this index. We show the scenario trees and present several statistical results as evidence for the quality of the scenario generation algorithm.

The thesis ends with concluding remarks.

Part I
Theoretical Foundation

2 Basic stochastic programming models and Scenario Trees

Investment and financial problems are characterized by decisions under uncertainty and over more than one period. To capture both these aspects, multi-stage (dynamic) stochastic programming seems to be the best framework in such a decisional environment.

The dynamic (recursive) nature of the program allows us not only to consider present available information, but also to embed future possible decisions in the optimization. Informally we could assert that with a dynamic model we try to define our best choice for today based on what we known today, and what we expect if we will make the best choice in the future, based on what we will know in the future. The stochastic nature, instead, allows us to consider the entire set of possible random variables (or at least an approximation of this set), and not just possible realizations (like the expected values).

To clarify these two concepts which seem, at least on a first reading a little bit fuzzy, in this section we describe how a stochastic program is formulated and solved. Two versions of a stochastic program are presented. We first describe the two-stage stochastic program with recourse and then the natural multi-stage extension. With the help of two simple examples, we explain the advantages of a dynamic stochastic program with respect to a single period deterministic optimization, like Markowitz's. We try to explain the dynamic nature of the program and why the optimal solution may be superior than for a static model. Finally a very simple financial planning problem is presented. This problem is similar to the asset allocation problem we want to solve. We will, therefore explain the random variable to consider and emphasize the importance of a restricted number of future scenarios, which have to describe properly future evolutions of the financial market.

2.1 Two-stage stochastic model with recourse

Recoursive (or dynamic) models are a combination of *anticipative models* and *adaptive models*. Anticipative models are also referred to as static models, for which the decision does not depend in any way on future observations of the environment. At the decision time we have to take into account all possible future realizations since the opportunity to adapt our decisions later on is not considered at the first-stage decision. In an adaptive model on the other hand, information related to uncertainty becomes partially available before decision making, thus optimization takes place in a learning environment, which is the essential difference with an adaptive model. The recoursive model combines these two models in a common mathematical framework, which seeks a policy that does not only anticipate future observations but also takes into account temporarily available information to make recourse decisions (Yu, Ji, and Wang (2003)).

In a financial framework, for example, a portfolio manager is always faced with recoursive models. For an optimal allocation of the assets in his portfolio, the manager considers both future movements of stock prices (anticipation) as well as the possibility of rebalancing the portfolio positions as prices change (adaption). Instead of considering both these aspects, the Markowitz static model limits itself in the anticipative expectation of future asset returns. It becomes, therefore, easy to guess why the Markowitz optimal results can be different than, and probably inferior to, the results obtained with a dynamic model.

Mathematically, the two stage stochastic models with recourse are:

$$\begin{aligned} \min \quad & Z(x) = f(x) + E_\omega \left[\mathcal{Q}(x, \omega)\right] \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{1}$$

where x is a n_1-dimensional vector of decision variables and represents the first-stage anticipative decision, which is made before the vector of random variables ω is observed, E_ω is the mathematical expectation with respect to ω, and $f(x) : R^{n_1} \rightarrow R$ is the first stage cost function and does not depend on ω. Corresponding to x are the first-stage constraint matrix and vector A and b of size $m_1 \times n_1$ and $m_1 \times 1$, respectively. In the second stage a number of random events $\omega \in \Omega$ may realize. $\mathcal{Q}(x, \omega)$ is the optimal value, for any given Ω, of the following nonlinear program

$$\begin{aligned} \min \quad & q(y, \omega) \\ \text{s.t.} \quad & Wy(\omega) = h(\omega) - T(\omega)x \\ & y \geq 0 \end{aligned} \tag{2}$$

where $q(y, \omega)$ is the second-stage cost function, and $y \in R^{n_2}$ is the second-stage recourse decision, which depends on the decision x taken in the first-stage, and on the random variable ω. The dependence of y on ω is of a completely different nature than the dependence of q or other parameters on ω. It is not functional but simply indicates that the decisions y are typically not the same under different realizations of ω. For a given realization of ω, the second-stage problem constraint matrix $T(\omega)$ and vector $h(\omega)$ become known, where $T(\omega)$ has size $m_2 \times n_1$ and $h(\omega)$, $m_2 \times 1$.

Note the dynamic nature of the recoursive model. In order to find the optimal first-stage decision x, we do not consider only the set of all possible future realisations Ω, but also the best second-stage decision y given all possible information we would have at stage two. To find the first-stage optimal solution, we work recursively, first solving the second-stage problem over the whole set of possible random variable realizations and the set of all possible first stage decisions, thus finding the first-stage optimal solution which maximizes the overall optimization problem. To capture the essence of such a dynamic stochastic optimization, and the advantages with respect to a static optimization repeated over time, consider the following two simple examples.

2.1.1 The taxi driving problem

A taxi driver has to go from node A to node G of a street network (see Figure 1). She may choose between different routes. At each intermediate node (street crossing, etc.), she can observe which (if any) of the following streets is congested. She can estimate how long it takes to drive a certain street segment if there is no congestion, and how long it takes if there is congestion. Moreover, she can estimate the probabilities of congestions.

Suppose the times in minutes to drive each street segment are like in Figure 1. Moreover, the only possible congested street is the one from point B to point E. If there is no congestion she needs 10 minutes to go from B to E, while with congestion the necessary

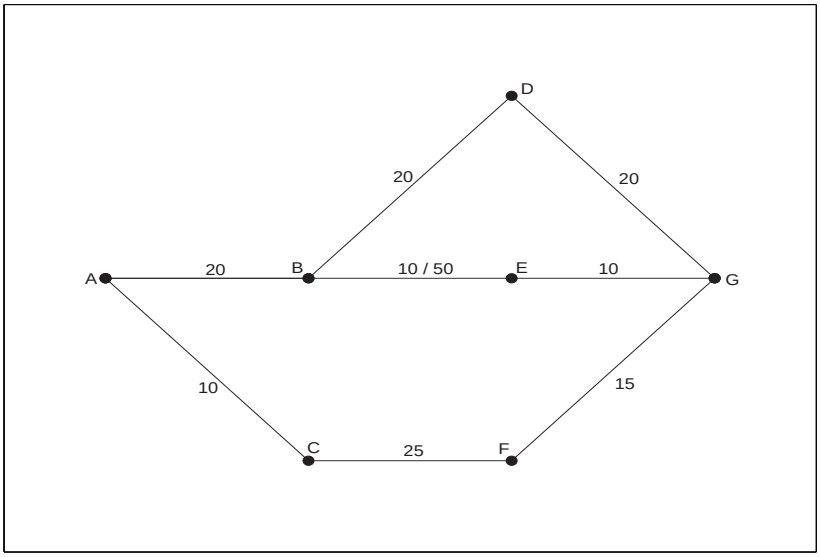

Figure 1: Example of a hypothetical street Network

time becomes 50 minutes. The probability of having a congestion is 30%. The decisions at stage one and stage two are $x^1 = \{B, C\}$ and $x^2 = \{D, E, F\}$, respectively. Obviously the second-stage decision depends on the decision taken at stage one.

Solving the problem with a static optimization model, we would choose the root with the minimal expected time:

$$\begin{aligned} E(ABDG) &= 20 + 20 + 20 = 60 \\ E(ABEG) &= 20 + 0.7 \cdot 10 + 0.3 \cdot 50 + 10 = 52 \\ E(ACFG) &= 10 + 25 + 15 = 50. \end{aligned}$$

Using a static (anticipative) optimization model, we would therefore choose the $ACFG$ root (since the expected distance time is minimal) and the first and second stage optimal decisions would be $x^1 = \{C\}$ and $x^2 = \{F\}$, respectively. This result however, does not consider the possibility that once having reached node B, to observe if there is a congestion on the street from B to E, or not. As mentioned above, under a static optimization we consider only the future expected values and not the information we could obtain in the future.

To solve the problem in a dynamic way, it is useful to build a decision tree (see Figure 2). With a white square we denote a decision node, with a circle a random event node while a black square represents an endnode (also called final leaf) of the decision tree.

Working backward from the endnodes we obtain that our first-stage optimal decision is to go from A to B and not from A to C like in the static solution. In fact, once we are at node B we can observe if the street from B to E is congested or not. If we are at node B and there is no congestion we choose the root BEG, with a time cost of 20. If, however, the route is congested, we choose the root BDG with a time cost of 40. The

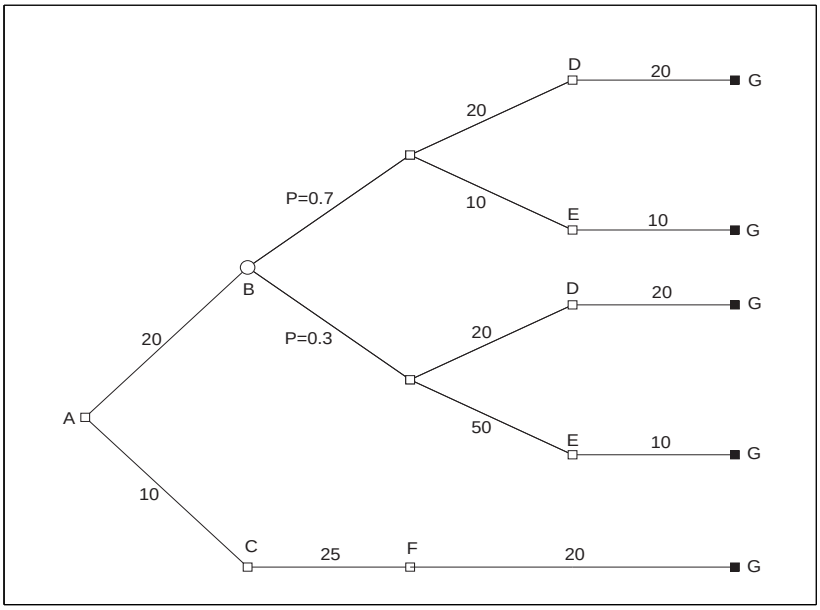

Figure 2: Decision tree for the Taxi driver example

expected time at node B is therefore $0.3 \cdot 40 + 0.7 \cdot 20 = 26$. Starting from A, we obtain an expected cost of $20 + 26 = 46$ going to node B and of $10 + 25 + 15 = 50$ going to node C. Our first-stage optimal decision is therefore to reach node B and not node C. Once point B is reached, we adapt our decision, conditioned by the information we obtain at stage two (is the street from B to E congested or not).

With a stochastic programming model we leave open the possibility of new information arriving in the system to react in an adaptive way. We leave the possibility open to change our decision in future periods, and moreover, we consider this possibility in the first-stage decision.

One could object that the example reported above is very particular and that in a portfolio selection problem such an extreme situation in which a decision taken today may limit future investment possibilities does not exist (at least not in the way as in the taxi driver example when street AC is chosen). This argumentation is valid. Two points are, however, to be considered. First, we have shown, although by a very specific example, that at least for some problems a stochastic program generates optimal solutions which are superior than those of a static program. Second, if we consider some liquidity constraints or the hedge funds markets, in which possible lock up or the decision of a manager to close his fund to further investments may restrict our future investment opportunities depending on our choice today, we could potentially find ourself in a situation similar to the one of the example.

The taxi driver example presented above considers a very restricted space of possible random events Ω, which consists of only two possible realizations. We can however solve a stochastic optimization model even if we have a continuous set of possible random

variables. The next example shows how to build and solve a continuous stochastic optimization problem. This example is used in the next section to analyze and understand several discretization procedures, necessary to restrict the probability set under which we work.

2.1.2 The news vendor problem

Let us suppose that every morning, before going to work we buy our newspaper from the boy working at the corner of our street. Knowing that we are economists, one day this news vendor asks us to help him in his everyday decision about how many papers he should buy from the publisher in order to maximize his everyday wealth. This problem is called the *'news vendor problem'* and is a classical example of stochastic programming. A simple version of the problem can be synthesized as follow:

Every morning a news vendor buys a quantity x of newspapers at a price of c per paper from a newspaper publisher. The vendor then sells as many newspapers as possible at a price p. The unsold newspapers cannot be returned to the publisher and are wasted. The question is how many newspapers the news vendor has to buy every morning in order to maximize his everyday wealth. Obviously the decision depends essentially on how many newspapers the news vendor can sell each day. If we call the daily demand for newspapers ω, and the quantity of sold newspapers $y(.)$, then the cost minimization problem (wealth maximization problem) is

$$\begin{aligned} &\min \{cx + E_\omega [-py(x, \omega)]\} \\ &\text{s.t.} \qquad x \geq 0 \end{aligned} \tag{3}$$

where E_ω denotes the mathematical expectation with respect to ω. Would the number of demanded newspapers be a known variable ω', constant, then the problem would be trivial (in order to maximize his wealth, the news vendor would buy a quantity of newspaper $x = y = \omega'$). If, more realistically, the demand for newspaper ω is not constant but is instead distributed with a density function $f(\omega)$ the problem would be more difficult but also more interesting:

The quantity of sold newspapers must follow the simple rule

$$y^*(\omega) = \min(\omega, x),$$

since the sales can never exceed the demand, nor the number of bought newspapers. The minimization problem can thus be rewritten as

$$Z(x) = cx + E_\omega [-p\ min(\omega, x)]\,. \tag{4}$$

$-E_\omega(.)$ is the expected profit on sales, while $py^*(\omega) = p\ min(\omega, x)$ is the profit on sales if the demand is at level ω. The model illustrates the two-stage aspect of the news vendor problem. The first-stage decision x is the quantity of newspapers that have to be bought from the publisher and the second-stage decision y is the number of newspapers that have to be sold. The buying decision has to be taken before any information is given about

the demand. The selling decision depends on the number of newspapers bought in stage one and on the demand for newspapers. When the demand becomes known in the second stage, the profit can be computed (Birge and Louveaux (1997), p. 16).

Solving equation (4) we obtain

$$\begin{aligned} Z(x) &= cx - p\int_{-\infty}^{\infty} min(\omega, x)\ f(\omega)d\omega \\ &= cx - p\int_{-\infty}^{x} \omega f(\omega)d\omega - p\int_{x}^{\infty} xf(\omega)d\omega \\ &= cx - p\int_{-\infty}^{x} \omega f(\omega)d\omega - px\int f(\omega)d\omega + px\int_{-\infty}^{x} f(\omega)d\omega \\ &= cx - px(1 - F(x)) - p\int_{-\infty}^{x} \omega\ f(\omega)d\omega \end{aligned} \tag{5}$$

where $F(x)$ is the cumulative distribution function at x. Integrating by parts, we obtain

$$\int_{-\infty}^{x} \omega\ f(\omega)d\omega = xF(x) - \int_{-\infty}^{x} F(\omega)d\omega$$

and

$$\begin{aligned} Z(x) &= (c-p)x + p\int_{-\infty}^{x} F(\omega)d\omega \\ &= (c-p)x + p\mathcal{F}(\omega) \end{aligned}$$

where

$$\mathcal{F}(\omega) = \int_{-\infty}^{x} F(\omega)d\omega.$$

Taking the first derivative,

$$Z'(x) = (c-p) + pF(x)$$

with the optimal solution

$$x^* = F^{-1}\left(\frac{p-c}{p}\right). \tag{6}$$

The result is intuitive. The optimal demand for newspapers is positively related to the selling price p and negatively related to the cost per paper c. The optimal result x^* depends essentially on the distribution F. It is, therefore, of primary importance to define the value and the probability of future random variables properly.

2.2 Multistage stochastic program

The recoursive problem is not restricted to the two-stage formulation. Most practical decision problems (like investment and financial problems) involve a sequence of decisions that react to outcomes that evolve over time. Observations are in this case made at T different stages and are captured in the information sets $\{I_t\}_{t=1}^{T}$. At stage τ the decision problem is conditioned on the information set I_τ which includes all the information in the sets $\{I_t\}_{t=1}^{\tau}$, and anticipates all the sets of future information $\{I_t\}_{t=\tau+1}^{T}$.

Let $y_1 \in R^{n_1}$ be the first-stage decision vector, and A and b the respective constraints. For each stage $t = 2, \ldots, T$, denote by $y_t \in R^{n_t}$, the future stages decision vectors, which depend on the random variable ω_t with support set Ω_t, and on the whole set of previously taken decisions. For each stage also define the random cost function $q_t(y_t, \omega_t)$, and the random parameters $\{T_t(\omega_t), h_t(\omega_t) | \omega_t \in \Omega_t\}$.

The multistage program, which extends the two-stage model of section 2.1, is formulated as the following nested optimization problem

$$
\begin{aligned}
\min \quad & Z = f(y_1) + E_{\omega_2}\left[\min_{y_2 \in R^{n_2}} q_2(y_2, \omega_2) + \ldots + E_{\omega_T}\left[\min_{y_T \in R^{n_T}} q_T(y_T, \omega_T)\right]\ldots\right] \\
\text{s.t.} \quad & Ay_1 = b \\
& W_2 y_2(\omega_2) = h_2(\omega_2) - T_2(\omega_2) y_1 \\
& W_3 y_3(\omega_3) = h_3(\omega_3) - T_3(\omega_3) y_2 \\
& \quad \vdots \\
& W_T y_T(\omega_T) = h_T(\omega_T) - T_T(\omega_T) y_{T-1} \\
& y_t \geq 0 \quad \text{for} \quad t = 1, \ldots, T
\end{aligned}
\tag{7}
$$

The increase of the number of stages in the optimization raises the difficulty of finding an optimal solution consequently. In the next example we present a very simplified version of an asset allocation problem over multiple periods. Although only three periods are considered and the portfolio assumes very particular values, the complexity of the optimization is still burdensome, and an optimization software is recommended to obtain an optimal solution.

2.2.1 Financial planning

Suppose that after many years working as a trader for one of the most important investment banks, we decided to give up our work and our beautiful apartment in the centre of London at the end of the year and to spend the rest of our life in a wonderful and microscopic atoll in the middle of the Pacific Ocean. For the beginning of our new life we have calculated a necessary amount of money equal to $\$G$. We currently have an amount of $\$B$ cash, which we want to invest in a portfolio of i assets, which we have to sell before leaving in order to gain the liquidity needed to cover our travel and living expenditures. Suppose we can rebalance the weights of our portfolio every υ^{th} month, so we have $D = 12/\upsilon$ investment periods.

This simple example ignores transaction costs and taxes on income although these considerations would be important in reality. The uncertainty of the problem is given by the return of each investment i within each period t. We describe this random variable as $\xi_i(t) = \xi_i(t, \omega)$ where ω is some underlying random element (See Birge and Louveaux (1997) for further considerations on a similar example).

For this simple example we assume that the portfolio consists of two assets only, a stock (i=s) and a bond (i=b). For every period we only have two independent scenarios S in which the return of the stock can be +27% (S=1) or -2% (S=2), while the return of the bond is 9% (S=1) or 5% (S=2). The movements of the stock and the bond are perfectly correlated (with correlation $\rho = 1$). In bullish markets, both assets gain the higher return (+27% for the stock and +9% for the bond) while in bearish markets the lower return is earned (-2% and 5%, respectively). Both scenarios have probability 1/2. Suppose that $\upsilon = 4$ months, so that the investment periods are $D = 3$. Figure 3 shows the scenario tree for this simple example.

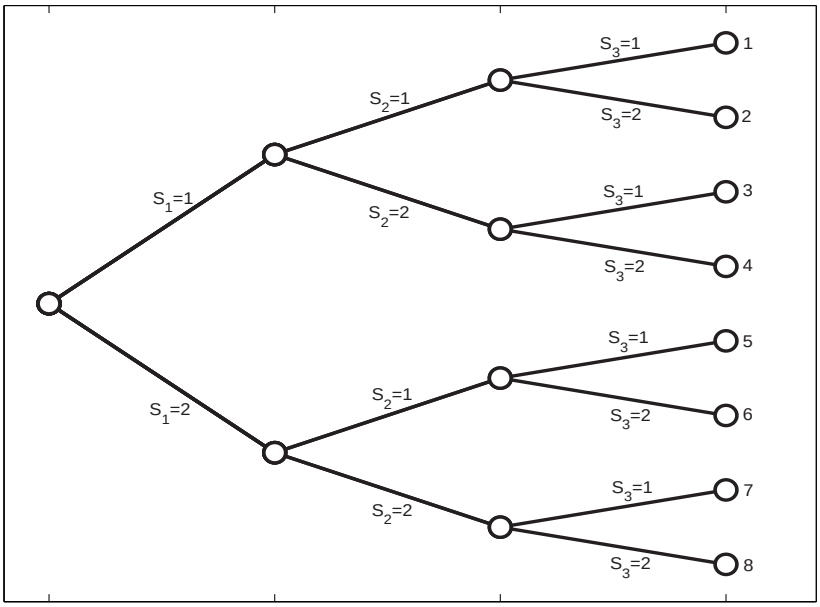

Figure 3: Structure of a simple three stages scenario tree with 2x2x2=8 scenario paths.

Since a final amount of money less than G would imply that we cannot pursue our dream, our primary goal is to avoid this possibility. For this reason, if we call y the surplus amount of money (G minus the final return of our portfolio > 0) and q the deficit (G minus the final return of our portfolio < 0), we can write the objective function to maximize as

$$\sum_{S_D} \dots \sum_{S_1} p(S_1, \dots, S_D)[(y(S_1, \dots, S_D) - Kq(S_1, \dots, S_D)]$$

where $p(S_1, \dots, S_D)$ is the probability of each of the possible scenarios, while K is the penalty we give to a deficit. In our simple example the probability of each scenario is equal to 0.125 ($= (1/2)^3$). The first-period constraint simply consists in investing the initial wealth

$$x_s(1) + x_b(1) = B.$$

The constraints for the intermediate periods (for t=2 to t=$D-1$) are, for each $S_1, \ldots, S_{t-1}$:

$$\sum_i -\xi_i(t-1, S_1, \ldots, S_{t-1}) x_i(t-1, S_1, \ldots, S_{t-2}) + \sum_i x_i(t, S_1, \ldots, S_{t-1}) = 0.$$

This intertemporal constraint states us that the amount of money gained in investment period $t-1$, is reinvested entirely in the assets in period t. The constraint for period D is

$$\sum_i \xi_i(D, S_1, \ldots, S_D) x_i(D, S_1, \ldots, S_{D-1}) - y(S_1, \ldots, S_D) + q(S_1, \ldots, S_D) = G.$$

The final investment gains, minus the surplus (or plus the deficit, depending on the investment output) has to be equal G by definition, the amount of money needed to attain the project. $\xi_i(t-1, S_1, ..., S_{t-1})$ is the return of asset i at time $t-1$ given that scenarios $S_1, ..., S_{t-1}$ occur, while $\xi_i(D, .)$ is the return of asset i in the last period $t = D$. $x_i(t-1, S_1, \ldots, S_{t-2})$ is the amount of money invested in asset i. This choice is made in period $t-1$, but is conditioned by the information at $t-2$, before the asset returns are revealed.

If B=55,000 is the initial wealth, and G=70,000 the target value, with a shortage penalty K=10, the stochastic program (in thousands of dollars) is

$$\begin{array}{lrcl}
\max_x & \left\{\sum_{S_1=1}^{2}\sum_{S_2=1}^{2}\sum_{S_3=1}^{2} 0.125(y(S_1,S_2,S_3)) - 10(w(S_1,S_2,S_3))\right\} & & \\
s.t. & x_s(1) + x_b(1) & = & 55, \\
& -1.27x_s(1) - 1.09x_b(1) + x_s(2,1) + x_b(2,1) & = & 0, \\
& -0.98x_s(1) - 1.05x_b(1) + x_s(2,2) + x_b(2,2) & = & 0, \\
& -1.27x_s(2,1) - 1.09x_b(2,1) + x_s(3,1,1) + x_b(3,1,1) & = & 0, \\
& -0.98x_s(2,1) - 1.05x_b(2,1) + x_s(3,1,2) + x_b(3,1,2) & = & 0, \\
& -1.27x_s(2,2) - 1.09x_b(2,2) + x_s(3,2,1) + x_b(3,2,1) & = & 0, \\
& -0.98x_s(2,2) - 1.05x_b(2,2) + x_s(3,2,2) + x_b(3,2,2) & = & 0, \\
& 1.27x_s(3,1,1) + 1.09x_b(3,1,1) - y(1,1,1) + w(1,1,1) & = & 70, \\
& 0.98x_s(3,1,1) + 1.05x_b(3,1,1) - y(1,1,2) + w(1,1,2) & = & 70, \\
& 1.27x_s(3,1,2) + 1.09x_b(3,1,2) - y(1,2,1) + w(1,2,1) & = & 70, \\
& 0.98x_s(3,1,2) + 1.05x_b(3,1,2) - y(1,2,2) + w(1,2,2) & = & 70, \\
& 1.27x_s(3,2,1) + 1.09x_b(3,2,1) - y(2,1,1) + w(2,1,1) & = & 70, \\
& 0.98x_s(3,2,1) + 1.05x_b(3,2,1) - y(2,1,2) + w(2,1,2) & = & 70, \\
& 1.27x_s(3,2,2) + 1.09x_b(3,2,2) - y(2,2,1) + w(2,2,1) & = & 70, \\
& 0.98x_s(3,2,2) + 1.05x_b(3,2,2) - y(2,2,2) + w(2,2,2) & = & 70, \\
& x_s(t, S_1, \ldots, S_{t-1}),\ x_b(t, S_1, \ldots, S_{t-1}) & \geq & 0, \\
& y(S_1, S_2, S_3),\ w(S_1, S_2, S_3) & \geq & 0, \\
& \text{for all } t,\ S_1,\ S_2,\ S_3. & &
\end{array}$$

The problem can be solved with an optimization program (we used GAMS[1]). The optimal solution (in thousand of dollars) appears in Table 1. In period 1 we invest heavily in stocks ($39.500) with only $15.500 in bonds. In period 2 we have the reaction to the first-period outcomes. After a "good" period (scenarios 1 to 4) we invest even more in stocks, while after a "bad" period (scenarios 5 to 8) the strategy is more conservative, with an increasing amount in the number of bonds. The last-period strategies are particularly interesting. In the best and worst scenarios (scenarios 1 to 2 and 7 to 8 respectively) we invest the whole amount in stocks. The reason is quite intuitive: If scenarios 1 or 2 occur, there is no risk of missing the target $\$G$, so that we can invest everything in the asset with the higher expected return (the stock). If we are in the worst scenario, the only hope to reach the target $\$G$ is to invest the whole amount in stocks, hoping for a bullish market in the next period. In scenarios 3 to 6, stocks may cause one to miss the target, so they are avoided and we invest only in bonds.

Period,Scenario	Stocks	Bonds
1, 1-8	39.5	15.5
2, 1-4	53.5	13.6
2, 5-8	37.4	17.6
3, 1-2	82.7	0.00
3, 3-4	0.00	66.7
3, 5-6	0.00	66.7
3, 7-8	55.1	0.00

Scenario	Above G	Below G
1	35.05	0.00
2	11.06	0.00
3	2.67	0.00
4	0.00	0.00
5	2.67	0.00
6	0.00	0.00
7	0.00	0.00
8	0.00	15.98

Table 1: Optimal solution with three period stochastic program.

The comparison between the results of the stochastic program of Table 1 and the results obtained with a deterministic model in which all random returns are replaced by their expected values is also interesting. Since in a deterministic model we only consider the expected values instead of all possible realizations of the random variables, we would invest the entire starting wealth B in the stock (since the expected return of the stock is equal to 1.125 while the one for the bond is 1.075). While with a stochastic optimization the value of the objective function equals -13.55, implementing a deterministic policy, under the same scenarios we obtain a loss function equal to -22.045. Choosing a linear

[1]The General Algebraic Modeling System (GAMS) is a high-level modeling system for mathematical programming and optimization. see www.GAMS.com

combination between the stock and the bond we obtain even worse results, since the overall expected value decreases. This comparison gives us a measure of the utility value in using a decision from a stochastic program compared to a decision from a deterministic optimization (Birge and Louveaux (1997) p. 25).

The reason for this positive value is first determined by a different description of the goal and the risk factors, as well as by the different number of periods considered in the optimization. In a static deterministic program during every period we maximize our expected wealth over a given level of variance. Given the perfect correlation between the returns of the stock and the bond, at every period all we do is to choose a linear combination between these two assets depending on our risk (or better, variance) aversion. In a stochastic program, we maximize over the whole considered investment horizon, and set a goal and a risk source, which better describe our needs (the attainment of a final level of wealth at least equal to G). It is therefore not surprising that the optimal solution of the dynamic stochastic program is sensibly better than the one obtained with a static deterministic model.

This example shows how a little increase of the problem complexity significantly increases also the difficulty of solving the optimization. In particular, increasing the number of considered periods, passing from a two-stage to a multistage optimization may cause the problem to be very difficult to solve. For more realistic examples, this is even worse. The number of nodes in the scenario tree increases exponentially with the number of considered periods. In a multistage optimization it is therefore easy to reach an unmanageable number of scenarios. This is because the problem becomes unsolvable or the computational time to obtain an optimal solution becomes extremely burdensome. For this reason we have to be very thrifty in the definition of future random variable realizations. Unfortunately, the optimal solution highly depends on the expectation of these realizations. A small change of the asset returns we expect for future periods may completely change the result of the optimization and our best investment strategy. On one side we cannot work with a high number of scenarios, on the other hand these scenarios have to describe the uncertainty we are faced with extremely well.

In the next section we show some statistical methods to reduce the number of scenarios given a well known density function. Supposing that we know exactly the continuous distribution of the random variables, we try to discretize them in such a way that the optimal solution remains unchanged or at least as close as possible to the solution under the unrestricted density function.

3 Scenario Discretization and Reduction

A key factor for the resolution of a stochastic program is the definition of the probability space, describing the possible realization and the occurrence probability of future random variables. The solution of the optimization highly depends on the choice of this probability space. According to Dupačová and Römisch (1998), unfortunately by the definition of the probability set, we are normally faced with incomplete information about the probability measure through which the stochastic program is formulated and also with the necessity of designing various discretization and approximation schemes in connection and evaluation of algorithms. In the next pages we deal with the second point, the discretization of the probability set.

In the financial planning example of section 2.2.1 we have seen how a little increase in the complexity of the problem increases according to the resolution difficulties and the computational time needed in the search for an optimum. For more realistic problems, these two aspects are even stressed, and obtaining an optimum may be impossible or the required time unacceptable for practical purposes. This is primarily due to the number of scenarios, which increase exponentially with the number of periods considered. In order to solve realistic problems (in our case a realistic asset allocation model), we have to restrict the considered probability space. This must be done very carefully, since the goal is to restrict the set of possible random variables trying to maintain the optimal solution unaltered.

In this section we present several techniques to reduce the probability space of possible random variables. We suppose that the information about the probability set is complete and that the continuous probability density function we want discretize is perfectly known. We start describing the discretization criteria and define the approximation error to minimize. We then present the different discretization techniques and try to identify if a superior methodology exists.

3.1 Discretization criteria and approximation error

Except for extremely simple and unrealistic cases, continuous-state multiperiod financial optimization problems can only be formulated, but not solved. The reason for practical insolvability is the fact that decisions are functions, making the problem a functional optimization problem, which cannot be numerically solved as it is (Pflug (2001)). The usual way of reducing the problem to a solvable one is to restrict it to a discrete-state multiperiod optimization problem. The random variable ω is reduced to a vector of finite values (i.e. $\omega_1, \ldots, \omega_S$) .

Consider the linear stochastic program

$$\begin{aligned} Z(x) = \; & cx + E_\omega f(\omega, x) \\ \text{s.t.} \quad & A(x) = b \\ & x \geq 0 \end{aligned} \tag{8}$$

with P is the probability measure on $(\Omega, \mathcal{B})$ with Ω (the set of possible outcomes ω) a closed subset of $\mathcal{R}^s$, $\mathcal{B}$ the borel σ-field relative to Ω and $x \in \mathcal{R}^n$ the decision vector. f is a function from $\Omega \times \mathcal{R}^n$ to $\mathcal{R}^1$ measurable with respect to ω and lowersemicontinuous and

convex with respect to x. E_P is the expectation with respect to the probability measure P. The minimization problem in (8) is

$$\begin{array}{lll} Z(x) & = & \int_\Omega f(\omega, x) dP(\omega). \\ x \in X & & \end{array} \tag{9}$$

We can eliminate the constraints by defining $X \subset \mathcal{R}^n$ the convex nonempty set of feasible first-stage decisions, while the linear deterministic part (cx) is embedded (with a little abuse of notation) in the function f. Consider now $\nu(P)$ the optimal value of the objective function and $S(P)$ the solution set of the minimization of equation (9).

$$\nu(P) := \inf \{E_P f(\omega, x) : x \in X\}$$

$$S(P) := \{x \in X : E_P f(\omega, x) = \nu(P)\} = argmin_{x \in X} E_P f(\omega, x).$$

Suppose for simplicity that the optimal solution is unique. The goal is to find an approximated probability measure $\tilde{P}$, such that the optimal solution $\tilde{x}^*$ under $\tilde{P}$ is as close as possible (if not equal) to the optimal solution x^* under the unrestricted probability measure P. We minimize the approximated objective value

$$\begin{array}{lll} \tilde{Z}(x) & = & \int_\Omega f(\omega, x) d\tilde{P}(\omega), \\ x \in X & & \end{array} \tag{10}$$

so that the approximation error $e(Z, \tilde{Z})$ defined as

$$e(Z, \tilde{Z}) := Z(argmin_x \tilde{Z}(x)) - Z(argmin_x Z(x)) \tag{11}$$

is minimized. In the extreme optimal case when the approximation error is equal to zero, we can work with a restricted probability space obtaining notwithstanding this the same optimal solution. Always $e(Z, \tilde{Z}) \geq 0$. Typically, the error e defined by (11) is difficult to calculate, since the term $argmin_x Z(x)$ cannot generally be computed. It is easier, to get an upper bound for it using the following lemma (e.g. Lemma 1, Pflug (2001)).

Lemma 1 (Pflug,2001)

$$e(Z, \tilde{Z}) \leq 2 \sup_x \left| Z(x) - \tilde{Z}(x) \right|.$$

Proof.

Let $x^* \in \text{argmin } Z$ and $\tilde{x}^* \in \text{argmin } \tilde{Z}$. Set $\epsilon = \sup_x \left| Z(x) - \tilde{Z}(x) \right|$.

Let $M = \{x : Z(x) \leq Z(x^*) + 2\epsilon\}$. Suppose that $\tilde{x}^* \notin M$. Then

$$Z(x^*) + 2\epsilon < Z(\tilde{x}^*) \leq \tilde{Z}(\tilde{x}^*) + \epsilon \leq \tilde{Z}(x^*) + \epsilon \leq Z(x^*) + 2\epsilon.$$

This contradiction establishes $\tilde{x}^* \in M$, i.e.

$$e(Z, \tilde{Z}) = Z(\tilde{x}^*) - Z(x^*) \leq 2\epsilon.$$

□

Lemma 1 asserts that instead of minimizing $e(Z, \tilde{Z})$ defined as the difference between the objective function evaluated at the restricted and unrestricted optimal point, we can minimize a distance between the objective function under the two probabilities measures, i.e.

$$\begin{aligned} e(Z, \tilde{Z}) &= \int_\Omega f(\omega, \tilde{x}^*) P(d\omega) - \int_\Omega f(\omega, x^*) P(d\omega) \\ &\leq 2 \sup_x \left| \int_\Omega f(\omega, x) P(d\omega) - \int_\Omega f(\omega, x) \tilde{P}(d\omega) \right|. \end{aligned} \tag{12}$$

The advantage is that we do not have to compute the optimal solution under both measures (remember that to compute the optimal solution under P is what we want to avoid, since it is difficult if not impossible). What we have to do is choose $\tilde{P}$ in such a way that the sup-distance between P and $\tilde{P}$ is small. Obviously $|Z(x) - \tilde{Z}(x)|$ depends essentially on the distance of the probability measures, and it is equal zero only if $dP(\omega) = d\tilde{P}(\omega)$.

There are many approaches to find $\tilde{P}$, such that the distance between the probability measures is minimized. We now review the most important methodologies published in the recent literature, and explain the advantages (and drawbacks) of each methodology. This is done with aid of the news vendor example. For the simple version presented in section 2.1.2 we know exactly the optimal solution under the unrestricted continuous probability space. We can therefore compare this solution with the optimal points obtained under the restricted probability spaces computed with the different approximation techniques.

3.2 Old methodologies

The standard approach for approximating a continuous distribution by a discrete distribution is the following: (1) divide the outcome region into intervals, (2) select a representing point in each interval, and (3) assign a probability to each point.

An example is the "bracket mean" method. The output region is divided into N equally probable intervals, the representative point is the mean of the corresponding interval, and each point has probability $1/N$. Miller and Rice (1983) however point out that "bracket mean" methods always underestimate the even moments and usually underestimate the odd moments of the original distributions. Consider the simple version of the news vendor problem presented in section 2.1.2. Suppose the demand for newspapers ω is normally distributed with mean zero and variance one (although the demand obviously has to be nonnegative, notice that the problem is invariant with respect to translation and we may as well assume that F is a normal $N(\mu, \sigma^2)$ distribution, for which the probability mass on the negative axis is arbitrarily small, e.g. Pflug (2001)). Suppose that the selling price p is equal to 6, that the cost c of a newspaper is equal to 1, and that we want to approximate the continuous distribution $F(\omega)$ with only ten points. The ten points "bracket mean" approximation of a standardnormal distribution generates the equally probable points

$x_1 = \Phi^{-1}(1/20) = -1.6449$; $x_2 = \Phi^{-1}(3/20) = -1.0364$; ...; $x_{10} = \Phi^{-1}(19/20) = 1.6449$ all with probability $P(x_1) = P(x_2) = ... = P(x_{10}) = 1/10$. Figure 4 shows the ten-points approximation for a $N(0,1)$ distribution. The shaded area has probability $1/10$ ($\Phi^{-1}(1/10) = -1.2816$).

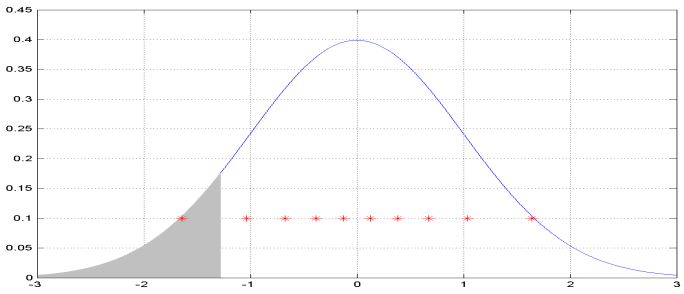

Figure 4: Ten-point "bracket mean" approximation of a $N(0,1)$ distribution.

With the given specifications the optimal solution of the objective function represented by equation (5) on page 12 is equal to

$$\begin{aligned} Z(x) &= x - 6x(1-\Phi(x)) - 6\int_{-\infty}^{x} \omega\ \phi(\omega)d\omega \\ &= x - 6x(1-\Phi(x)) - 6\int_{-\infty}^{x} \frac{\omega}{\sqrt{2\pi}} exp(-\omega^2/2) \\ &= x - 6x(1-\Phi(x)) + \frac{6}{\sqrt{2\pi}} exp(-x^2/2) \end{aligned} \tag{13}$$

which can be graphically represented by Figure 5 (see also Pflug (2001), Figure 1).

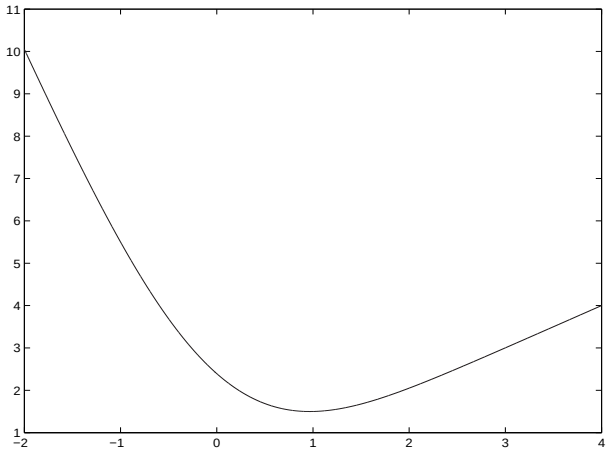

Figure 5: True objective function Z of the news vendor problem

The optimal solution for the news vendor problem given that the demand for newspapers is normally distributed with mean zero and variance one, and a buy and sell price of six and one respectively, can be computed by equation (6) on page 12, and is equal to:

$$x^* = \Phi^{-1}\left(\frac{p-c}{p}\right) = \Phi^{-1}\left(\frac{5}{6}\right) = 0.9674.$$

Consistently with Miller and Rice (1983), using the "bracket mean" method, we underestimate the second moment which is 0.88 instead of 1. The objective function $\tilde{Z}$ under the "bracket mean" approximation, looks as in Figure 6. Inserting the ten possible scenarios in equation (13) we obtain the optimal solution for $\tilde{Z}$ equal to $\tilde{x}^* = \Phi^{-1}(15/20) = 0.675$.

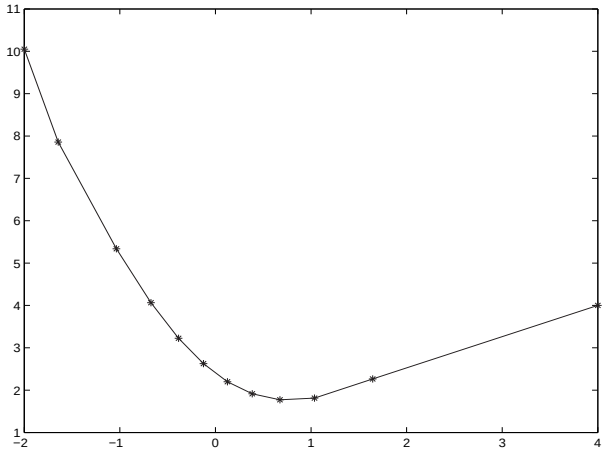

Figure 6: "Bracket mean" approximated objective function $\tilde{Z}$ of the news vendor problem

The distance between the optimal solution obtained with the continuous density function x^* and with the approximated function $\tilde{x}^*$ is 0.2924, while the approximation error $e(Z, \tilde{Z})$ equals 0.0701.

3.3 Moment matching techniques

Another class of approximation techniques consists in the preservation of the moments of the original distribution. Miller and Rice (1983) compute a discrete approximation of probability distributions by using the Gaussian quadrature method of numerical integration. This approach approximates the integral of the product of a function $g(x)$ and a weighted function $w(x)$ by evaluating $g(x)$ at several values of x, and computing a weighted sum of the results:

$$\int_a^b g(x)w(x)dx = \sum_{i=1}^{N} w_i g(x_i).$$

To establish the correspondence between the numerical integration formula and a discrete approximation of a probability distribution, the authors associate the distribution $g(x)$, with the weighted function, $w(x)$, and the probabilities, p_i, with the weights, w_i. They approximate $g(x)$ by a polynomial, and choose x_i and p_i (or w_i) to provide an equation for each term of the polynomial. The idea is to find x_i and p_i such that the first $2N-1$ moments are perfectly matched, i.e. so to satisfy the following equations:

$$
\begin{array}{ccccccccc}
p_1 & + & p_2 & + & \ldots & + & p_N & = & E(x^0) = 1, \\
p_1 x_1 & + & p_2 x_2 & + & \ldots & + & p_N x_N & = & E(x), \\
p_1 x_1^2 & + & p_2 x_2^2 & + & \ldots & + & p_N x_N^2 & = & E(x^2), \\
\vdots & & & & & & & & \\
p_1 x_1^{2N-1} & + & p_2 x_2^{2N-1} & + & \ldots & + & p_N x_N^{2N-1} & = & E(x^{2N-1}).
\end{array}
$$

To solve these equations the authors first construct the polynomial

$$\pi(x) = (x - x_1)(x - x_2)\ldots(x - x_N) = \sum_{k=0}^{N} C_k x^k$$

with $C_N = 1$ and $\pi(x_i) = 0$ for $i = 1, 2, \ldots, N$. Manipulating the first set of equations with the polynomial $\pi(x)$, it is possible to yield the following set of linear equations:

$$
\begin{array}{ccccccccc}
E(x^0)C_0 & + & E(x)C_1 & + & \ldots & + & E(x^{N-1})C_{N-1} & = & -E(x^N), \\
E(x)C_0 & + & E(x^2)C_1 & + & \ldots & + & E(x^N)C_{N-1} & = & -E(x^{N+1}), \\
E(x^2)C_0 & + & E(x^3)C_1 & + & \ldots & + & E(x^{N+1})C_{N-1} & = & -E(x^{N+2}), \\
 & & \vdots & & & & & & \\
E(x^{N-1})C_0 & + & E(x^N)C_1 & + & \ldots & + & E(x^{2N-1})C_{N-1} & = & -E(x^{2N-1}).
\end{array}
$$

The algorithm to compute the x_i and the p_i consists first in solving the last set of N equations in order to find the coefficient of the polynomial $C_0, \ldots, C_{N-1}$. Second, in finding the x_i computing the roots of the polynomial $\pi(x)$ and finally in determining the p_i by substituting the x_i into the original set of equations for the moments of the approximated distribution.

Consider again the news vendor problem of section 2.1.2. The approximation based on the Gaussian integration rules of Miller and Rice (1983) generates the following 10 points:

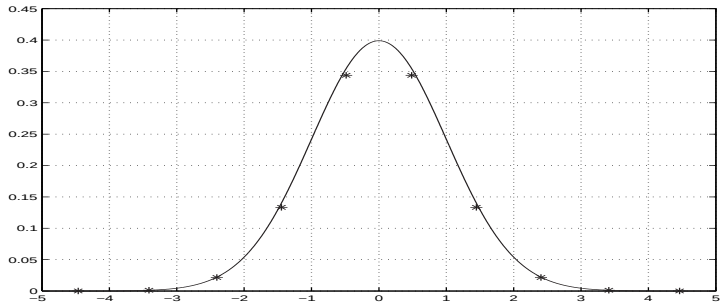

Figure 7: Ten-point approximation of a $N(0, 1)$ distribution, using a Gaussian quadrature method

It can be graphically noted that the ten-point Gaussian quadrature approximation fits the slope of the probability density function very well. In particular comparing the approximation points with those obtained with a "bracket mean" approximation, it seems that the Miller and Rice (1983) technique describes the function better. This, however, does not guarantee a better solution of the minimization problem. Even if the first 19 moments $(2N - 1)$ are perfectly matched, and the original distribution seems to be, at

least graphically quite good described by the approximation, if we insert the ten possible scenarios in the loss function of equation (13), what we obtain is a worse solution than the one obtained with the "bracket mean" method. The approximated loss function is shown in Figure 8:

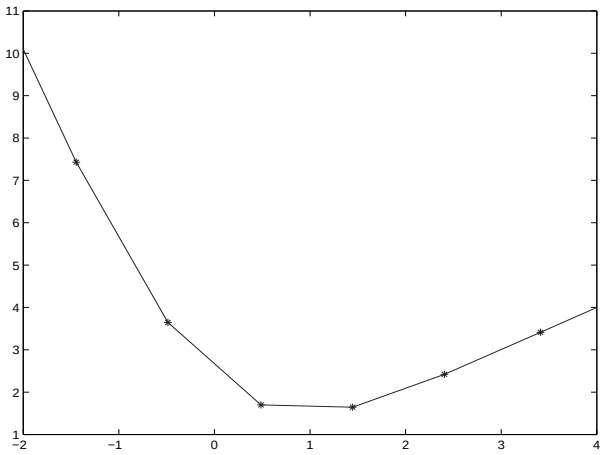

Figure 8: Approximated objective function $\tilde{Z}$ of the news vendor problem using a Gaussian quadrature method

The optimal solution which minimizes $\tilde{Z}$ is equal $\tilde{x}^* = 1.445$. The distance from the optimal solution using the entire original distribution equals 1.466-0.9674=0.4986 (while using the "bracket mean" approximation, this distance was equal to 0.2924). The approximation error is $e(Z, \tilde{Z}) = 0.1456$ (while it was equal to 0.0701 for the "bracket mean" method).

These results seem to suggest that the "bracket mean" method is better than the algorithm proposed by Miller and Rice (1983). As you can easily imagine, however, this result cannot be generalized and is very specific to this particular example. Changing the number of approximation points from ten to eleven, for example, gives an opposite result. The approximation error for the Gaussian quadrature is in this case equal to $e = 0.010$, while for the "bracket mean" it is $e = 0.012$. Also changing the parameters of the problem would change the quality of the different approximations. This shows how difficult it is to define a general criterion to choose which discretization method is better in describing the original distribution and computing the optimal solution for the minimization problem.

Another problem of the Miller and Rice (1983) method, is the limited number of scenarios we can compute. This limit is due to the fact that we have to calculate the roots of the polynomial $\pi(x)$, which limits the number of x_i to around 15-20 points if the moments do not explode too rapidly. In the case of distributions with a high growth of their central moments, the number of mass points is even reduced. For instance, we will see later on that if we discretize an exponential distribution, the maximal number of approximation points has to be reduced to 6. Moreover, it can be the case that for some densities the central moments are undefined or equal infinity. For a student-t distribution, often used to describe assets returns, only the first five moments are defined and less than infinity. For this distribution the Miller and Rice method allows only a 3 points approximation. Besides

the limited number of computable scenarios, in the case of a multiperiod problem, like the financial planning problem of section 2.2.1, or a multidimensional continuous distribution, due to multiple risk factors, it would be suboptimal to consider every density function separately, since specifying the distribution only locally to each node or to each risk factor would cause a lack in direct control of statistical properties defined "overall" outcomes in the later periods or of statistical properties connecting all different distributions in the problem (see N., Rustem, and Settergren (2001)).

To avoid these problems, Høyland and Wallace (2001) propose another algorithm to generate the desired scenarios (a more recent and ameliorated version of the algorithm is presented in Høyland, Kaut, and Wallace (2003)). In order to discretize a distribution, the authors minimize a "squared distance of statistical properties". In particular they construct the two vectors of x and p (the values and the probabilities of the discrete distribution) so that the statistical properties of the approximating distribution match the specified statistical properties, subject to constraints defining the probabilities to be nonnegative and to sum up to one. For a simple one-period decision model like the news vendor problem, the minimization can be stated as follows:

$$\begin{aligned} \min_{x,p} \sum_{i \in S} (f_i(x,p) &- S_{VAL_i})^2 \\ \sum p &= 1 \\ p &\geq 0. \end{aligned} \tag{14}$$

Where S is the set of all specified statistical properties, and S_{VAL_i} is the value of statistical property i in S, while $f_i(x,p)$ is the statistical property i of the approximated distribution. In particular, the authors use the first four central moments as statistical property, i.e. Equation (14) minimizes the sum of the squared difference of the first four central moments.

A. First set of starting points				B. Second set of starting points			
S. Points	x	p	$Z(x)$	S. Points	x	p	$Z(x)$
-2.3	-3.962	0.0003	19.8101	-3.565	-4.574	0.0001	22.87
-1.5	-2.370	0.033	11.8679	-2.247	-3.081	0.006	15.4067
-0.8	-1.115	0.242	5.9747	-1.035	-1.624	0.144	8.2517
-0.4	-0.189	0.096	2.8143	0.133	-0.322	0.482	3.1607
-0.1	-0.182	0.128	2.7972	1.319	**0.944**	**0.321**	**1.4995**
0.1	-0.115	0.077	2.6395	1.908	1.977	0.026	2.0312
0.4	**0.637**	**0.314**	**1.5895**	2.393	2.181	0.0004	2.2119
0.8	1.764	0.105	1.8577	2.936	2.435	0.014	2.4497
1.5	2.923	0.002	2.9260	3.696	2.691	0.005	2.6976
2.3	3.235	0.003	3.2360	5.469	4.340	0.0001	4.3400

Table 2: Ten-point Moment Matching approximation of a $N(0,1)$ distribution.

In the case of multiple continuous distributions, they add the correlations in the statis-

tical properties to be minimized, in order to consider the connection between the different distributions.

The main drawback of this methodology is given by the non convexity of the minimization problem (14). This possibly prevents us from reaching a global minimum easily and makes our problem very dependent on the starting points (needed in the case of a nonlinear optimization problem). If, for instance, we use the first four central moments and the correlation between the random variables as loss function to minimize, we would find different (sub-)optimal solutions depending on the starting points. Table 2 shows, for example, the results obtained for the news vendor problem under the stated specifications and with two different sets of starting points. Both discrete distributions approximate well the original $N(0,1)$ density well enough, matching the first four central moments perfectly. For both sets the minimum value of the loss function in equation (14) is of order $E-10$, suggesting the attainment of two different local minima.

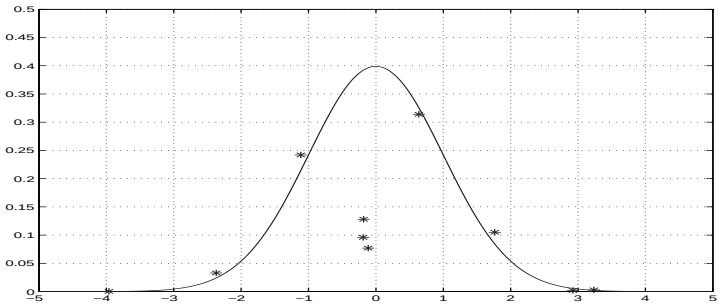

Figure 9: Ten-point "Moment Matching" approximation of a $N(0,1)$ distribution, with starting points reported on Table 2A.

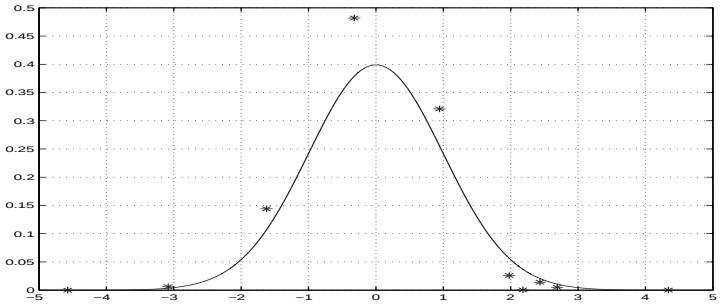

Figure 10: Ten-point "Moment Matching" approximation of a $N(0,1)$ distribution, with starting points reported on Table 2B.

Obviously, the optimal solutions of the objective function (13) are also different. The optimal value using the first set of starting points is equal $\tilde{x}_1^* = 0.637$, while the one using the second set is equal $\tilde{x}_2^* = 0.944$. Figure 9 and 10 represent graphically the distribution approximation. Note that the two approximations are quite different.

To avoid different optimal solutions (in the case for example of a unique random variable) we should consider not only the first four central moments (or a low number of statistical properties) but instead a higher number of constraints. Increasing the number of statistical properties to be matched, we decrease the number of degrees of freedom by

reducing automatically the influence of the starting values in the computation of an optimal discretization. In the extreme case, where we match $2N-1$ moments, we are able to reach a unique solution equal to the one found by Miller and Rice (1983). This partially solves the problem of the limited number of scenarios using the Gaussian quadrature method.

Also in this case, however, the perfect match of the first four moments does not guarantee a good optimal value. Using the two different sets of starting points, the approximation error is $e = 0.0904$ and $e = 0.0004$ respectively. For the first set the approximation error is even higher than the one obtained with the "bracket mean" method. These results are consistent with the conclusions drawn by Keefer (1994): Even if the approximation estimates the first moments of an underlying distribution exactly, the expected utility or the certainty equivalent of that distribution will not necessarily be well-approximated. This suggests the *possibility* that approximations, which have been recommended primarily based on their abilities to estimate mean, variance and other moments, may not perform well in estimating expected utilities or certainty equivalents. For this reason, other authors like Pflug (2001), Hochreiter and Pflug (2003), Dupačová, Gröwe-Kuska, and Römisch (2002) among others use probability metrics to find the optimal approximation.

3.4 Probability metrics techniques

Consider again the approximation error, $e(Z, \tilde{Z})$ defined in equation (11),

$$e(Z, \tilde{Z}) := Z(argmin_x(\tilde{Z}(x)) - Z(argmin_x(Z(x))$$

and the following definition of a *transportation metric* (see Pflug (2001) definition 1).

Definition 1 (Pflug, 2001)
Let $L_1(f)$ be the Lipschitz-constant of f, i.e.

$$L_1(f) = inf\left\{L : |f(u) - f(v)| \leq L\,|u - v| \ \forall\ u, v\right\} \tag{15}$$

(if there is no such L_1, we set $L_1(f) = \infty$). The Wasserstein distance d_W between P and $\tilde{P}$ is defined as

$$d_W(P, \tilde{P}) = sup\left\{\int f(u)dP(u) - \int f(u)d\tilde{P}(u) : L_1(f) \leq 1\right\} \tag{16}$$

Equation (16) is similar to the sup-distance we want to minimize, reported in equation (12). We see later on that to minimize the Wasserstein distance is equivalent to minimizing this sup-distance.

The Kantorovich-Rubinstein Theorem states that the Wasserstein distance may equivalently be defined as the infimum of the expected value of the geometric distance between two random variables jointly distributed with marginal distribution function P and $\tilde{P}$. Moreover Pflug (2001) reports the following properties for the Wasserstein distance:

Theorem 1
The d_W distance has the following properties:

1. *(Kantorovich-Rubinstein)*

$$d_W(P,\tilde{P}) = \inf\{E(\|X-Y\| : (X,Y) \text{ is a bivariate r.v. with given marginal distribution function } X\sim P; Y\sim \tilde{P}\}.$$

2.

$$d_W(P,\tilde{P}) = \int \left|P(u)-\tilde{P}(u)\right| du = \int \left|P^{-1}(u)-\tilde{P}^{-1}(u)\right| du$$

Where $P^{-1}(u) = sup\{v : P(v) < u\}$.

3. *Among all* $\tilde{P}$*, which sit on the mass point* $z_1, z_2, ..., z_k$*, the one closest to* P *in* d_W*-distance is*

$$\tilde{P}(x) = \sum_{\{i:z_i\le x\}} P\left(\frac{z_i+z_{i+1}}{2}\right)$$

where $z_k+1=\infty$ *and* $P(\infty)=1$*. For this* $\tilde{P}$*, the supremum in (16) is attained by*

$$f(u) = \min_i |u-z_i|,$$

i.e.

$$d_W(P,\tilde{P}) = \sum_{i=1}^{k} \int_{\frac{z_{i-1}+z_i}{2}}^{\frac{z_i+z_{i+1}}{2}} |u-z_i|\, dP(u)$$

where $z_0 = -\infty$.

Proof. For (1) see Rachev (1991), Theorem 5.3.2 and Theorem 6.1.1. (2) is proved in Vallander (1973). (3) is an easy consequence of (2). □

Part 3 of Theorem 1, tells us how to find the mass points $z_1, ..., z_k$ and the respective probabilities in order to optimally approximate a continuous distribution with respect to the Wasserstein distance. First compute the mass points that minimize the supremum of the Wasserstein distance in (16), i.e.

$$min\left\{\int \min_i |u-z_i|\, dP(u) : z_1, ..., z_k \in R\right\}$$

then compute the respective probabilities, determined by (see Hochreiter and Pflug (2003))

$$p_i = \int_{\{u:|u-z_i|=\min_j|u-c_j|\}} d(P(u)).$$

Suppose that the cost functions $\omega \to f(\omega, x)$ are uniformly Lipschitz, i.e. for all $x \in X$, $L_1(f) = inf\{L : |f(u)-f(v)| \le L|u-v|\} \le \bar{L}_1$, then

$$\sup_x \left|Z(x)-\tilde{Z}(x)\right| \le \bar{L}_1 \cdot d_1(P,\tilde{P}).$$

The original problem of minimizing $\sup_x \left|Z(x) - \tilde{Z}(x)\right|$ can be approximated by the problem of minimizing the Wasserstein distance, which is equivalent to finding the discrete $\tilde{P}$, which is closed to P in the mass transportation sense.

Consider again the news vendor problem. Suppose that we want to discretize the continuous distribution of the demand of newspapers with ten mass points. Because of the symmetry of the normal distribution, the guess about the points we use to discretize the distribution is $-5z, -4z, -3z, -2z, -z, z, 2z, 3z, 4z, 5z$. We now compute the Wasserstein distance and find the z which minimizes d_W. Using Theorem 1, and after some tedious algebra, the distance d_W is

$$\begin{aligned} d_W(\Phi, \tilde{P}_z) &= \frac{4}{\sqrt{2\pi}}\left[exp\{-\frac{25}{2}z^2\} - exp\{-\frac{81}{8}z^2\} + exp\{-8z^2\} - exp\{-\frac{49}{8}z^2\}\right. \\ &\quad \left. + exp\{-\frac{9}{2}z^2\} - exp\{-\frac{25}{8}z^2\} + exp\{-2z^2\} - exp\{-\frac{9}{8}z^2\} + exp\{-\frac{1}{2}z^2\}\right] \\ &\quad + 20z\Phi(5z) - 18z\Phi(\frac{9}{2}z) + 16z\Phi(4z) - 14z\Phi(\frac{7}{2}z) + 12z\Phi(3z) \\ &\quad - 10z\Phi(\frac{5}{2}z) + 8z\Phi(2z) - 6z\Phi(\frac{3}{2}z) + 4z\Phi(z) - 11z - \frac{2}{\sqrt{2\pi}}. \end{aligned}$$

Figure 11 shows graphically the distance d_W with respect to different values of z. The z which minimizes the Wasserstein distance is $z = 0.3406$.

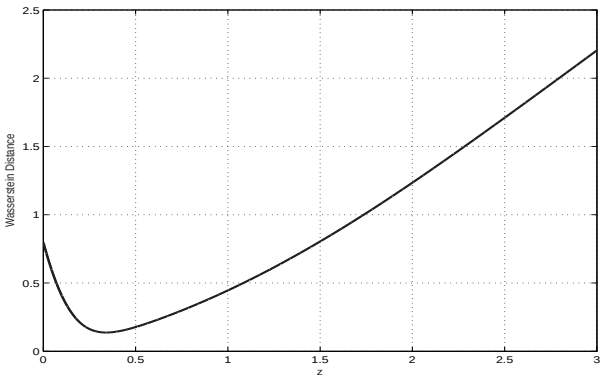

Figure 11: The d_W distance between Φ and $\tilde{P}_z$ for $0 \leq z \leq 3$

We thus find the ten points and the respective probabilities

value	± 1.703	± 1.3624	± 1.0218	± 0.6812	± 0.3406
probability	0.0627	0.0539	0.0806	0.1075	0.1953

which are graphically represented in Figure 12. With this approximation, the optimal value of x which minimizes the objective function Zx) of equation (9) is equal $\tilde{x}^* = 1.0218$. The approximation error is equal $e(Z, \tilde{Z}) = 0.0022$. The approximation error is second only to the error obtained with the moment matching method proposed by Høyland and Wallace (2001), using the second set of starting points. Considering, however, the

relative high number of degrees of freedom and the related wide spectrum of possible solutions using only the first four central moments, it seems that the Wasserstein distance minimizations effectively result in a better approximation (at least for the points near the optimal point) than the other approximation methods. As seen before, however, we must be very careful in judging the quality of a discretization methodology. It could be possible that the d_W-distance works well for this specific example but not for similar problems with different parameters, or a different number of approximation points or finally for different distributions.

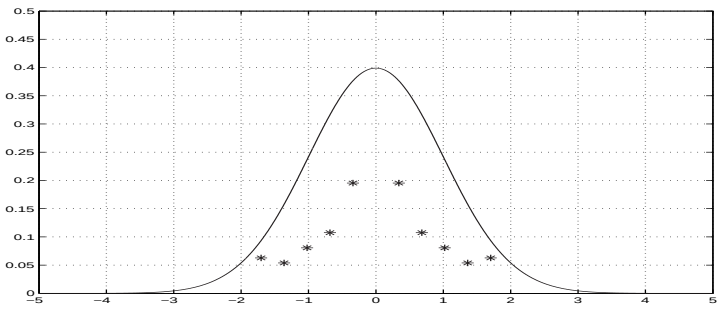

Figure 12: Ten-point approximation of a $N(0,1)$ distribution, under the Wasserstein distance minimization

3.5 Does a better approximation method exist?

As seen in the previous section, it is probably impossible to determine analytically whether an approximation method works better than others. The results are very problem specific to the considered problem and, depending on the specifications and the number of mass points we want, we draw different conclusions. We will, however, try in this exercise, analyzing different density functions and different numbers of mass points. In particular, we will again consider the news vendor problem, but instead of assuming that the demand for newspapers is $N(0,1)$, we also consider an exponential distribution

$$\psi(x) = \lambda e^{-\lambda x},$$

with $\lambda = 1$, which is well known to be non-symmetric, and a student-t distribution

$$\theta(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1 - \frac{x^2}{\nu}\right)^2},$$

with $\nu = 5$, which has been shown to be a good distribution for the description of assets returns.

For the $N(0,1)$ case with the cost for one newspaper being $c = 1$ and the price being $p = 6$, we know the optimal solution is: $x^* = \Phi^{-1}(5/6) = 0.9674$. To compute the optimal solution for the exponential case, consider again the cost minimization problem treated in section 2.2.1:

$$Z(x) = cx + E_\omega \left[-p \min(\omega, x)\right],$$

where again, $-E_\omega(,)$ is the expected profit on sales, $p\ \min(\omega, x)$ is the profit on sales if the demand is at level ω, and cx are the costs of buying the quantity x of newspapers. Solving this cost-minimization problem yields:

$$\begin{aligned} Z(x) &= cx - p\int_{-\infty}^{\infty} min(\omega, x)\ \psi(\omega)d\omega \\ &= cx - p\int_{0}^{x} \omega\psi(\omega)d\xi - p\int_{x}^{\infty} x\psi(\omega)d\omega \\ &= cx + p(xe^{(-x)} + e^{(-x)} - 1) - px(1 - \Psi(x)) \end{aligned} \tag{17}$$

where $\psi(\omega)$ is the probability density function of the exponential distribution with $\lambda = 1$, i.e. $f(\omega) = e^{(-\omega)}$, while $\Psi(x)$ is the cumulative distribution function evaluated at x and equal to $\Psi(x) = 1 - e^{(-x)}$. Under the given specification, equation (17) is graphically represented in Figure 13. Taking first derivatives, we obtain

$$\frac{\partial Z(x)}{\partial x} = c - pe^{(-x)} \tag{18}$$

with the minimum equal to:

$$x^* = \ln\left(\frac{p}{c}\right) = \ln(6) = 1.79.$$

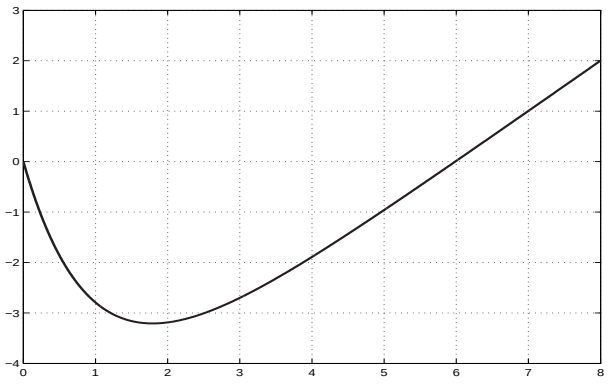

Figure 13: Objective function Z of the news vendor problem with the demand for newspapers exponentially distributed with $\lambda = 1$

The optimal solution of the news vendor problem under a student-t distribution is instead, equal to

$$x^* = \Theta^{-1}\left(\frac{p-c}{c}\right) = \Theta^{-1}\left(\frac{5}{6}\right) = 1.075$$

where Θ represents the student-t cumulative distribution function.

For every one of the four discretization methods analyzed previously, we discretize the three distributions with three to nine mass points. The results appear in Tables 4 to 9. In

bold are the points which minimize the respective loss functions. In order to also consider the found probabilities, we use them to compute the cumulative density function, as if we did not know the functional form of the original density, but only the resulting scenarios. For each discretization we also report the approximation error $e(Z, \tilde{Z})$, and we mark with a star the point among the different methods which minimizes this error.

For the exponential and the student-t distributions, the Gaussian quadrature method allows finding maximally a six- and a three-point approximation, respectively. For the exponential distribution the reason is to be found in the rapid growth of the values of the moments of this distribution, which makes it impossible to compute the roots for polynomials of degree higher than six (this problem also appears using the distance minimization technique of Høyland and Wallace (2001) if more than 11 central moments are considered in the minimization). For the student-t distribution, instead, only the first five central moments are defined. All other moments are in fact undefined (the odd moments) or equal infinity (the even moments). These two examples point out an important drawback of the discretization procedures based on central moments.

Second, for the method proposed by Høyland and Wallace (2001) (called "distance minimization") we minimize the sum of the squared difference of the first four central moments. Since the results for this methodology depend essentially on the starting points, and because of the similitude with the Gaussian quadrature method of Miller and Rice (1983), we use the mass function obtained from this last method as starting points, and then change them (in the case some scenarios have probability equal zero) until the desired number of different scenarios is found. In those cases for which it is impossible to compute the mass points with a Gaussian quadrature method, we use the mass function resulting from the Wasserstein minimization criteria as starting values.

Although the results do not show a clear advantage in using one approximation method instead of another, several conclusions can still be made. First consider the number of times an approximation reaches the minimal error (i.e. it is the best approximation among all different methods). As summarized in Table 3, for the normal and the exponential distribution the maximal number of 'best approximations' is reached by the Wasserstein distance method, while for the student-t distribution the best method seems instead to be the one proposed by Høyland and Wallace. Not surprising is the difficulty of the "bracket mean" method in describing very skewed distributions, like the exponential one.

Second, if we characterize as *'good'* and *'very good'* a discretization procedure implying an approximation error less then 0.05 and 0.01 respectively, we note that the Wasserstein distance minimization beats every other discretization method. In particular for the exponential and the student-t distribution for almost every approximation (5 for the exponential and 6 for the student-t density) the approximation error is less than 0.05. The number of *'good'* and *'very good'* discretizations is drastically reduced for the other three methods, in particular for the "bracket mean" and the Gaussian quadrature method. Characterizing instead as *'bad'* a discretization which involve an approximation error greater than 0.2 similar conclusions are drawn. For all three distributions, the Wasserstein distance minimization criteria generates only one approximation greater than 0.2. In fact, this happens only for the three point discretizations. All other discretization methods (in particular the "bracket mean") show a higher number of *'bad'* approximations.

To summarize the Gaussian quadrature method proposed by Miller and Rice (1983) cannot be computed when the central moments of the original distribution grow too rapidly, are not defined or equal infinity. The "bracket mean" method works quite badly for every one of the analyzed distributions. Especially for asymmetric distributions this method is beat by all other discretization procedures and, in general, the number of approximation error greater than 0.2 is higher than for the other methods. Finally, the Høyland and Wallace discretization method seems to be the best in discretizing a student-t distribution. Unfortunately, this method is very sensitive to the starting values, and it often happens that one or more mass points end up having a probability equal to zero. This causes the discretization to be very tedious, because the starting values have to be changed several times until the desired number of mass points is reached.

The Wasserstein method proposed by Pflug (2001) among others, seems the more stable discretization method. For any of the analyzed distributions the value of the approximation error is small and if we do not consider the three point discretization, never higher than 0.2. Moreover, asymmetry or fat tails seem not to play any important role in the quality of this discretization. In fact, the Wasserstein method works also very well with a high skewed distribution like the exponential one or with a leptokurtic distribution like the student-t. For all these reasons we decide to use for our scenario generation algorithm a Wasserstein distance minimization criteria.

Normal distribution: N(0,1)

	Bracket Mean	Gaussian Quadrature	Distance Minimization	Wasserstein-distance
Best	1	2	1	2
$e < 0.01$	1	1	2	2
$e < 0.05$	2	3	3	3
$e > 0.20$	2	2	3	1

Exponential distribution: Exp(1)

	Bracket Mean	Gaussian Quadrature	Distance Minimization	Wasserstein-distance
Best	0	1	2	4
$e < 0.01$	0	1	0	2
$e < 0.05$	1	1	2	6
$e > 0.20$	4	0	2	1

Student-t distribution: Std(5)

	Bracket Mean	Gaussian Quadrature	Distance Minimization	Wasserstein-distance
Best	2	0	3	2
$e < 0.01$	1	0	0	3
$e < 0.05$	2	0	4	5
$e > 0.20$	4	1	1	1

Table 3: Statistics of different approximation methods in approximating a Normal and an Exponential distribution with different numbers of mass points

	"Bracket Mean"		Gaussian Quadrature		"Distance Minimization"		Wasserstein-distance	
3	-0.9674	0.333	-1.7321	0.1667	-1.736	0.166	-1.029	0.3035
	0.00	**0.333**	0.00	0.667	-0.004	0.667	**0.00**	**0.3930**
	0.9674	0.333	**1.7321**	**0.1667**	**1.728**	**0.168***	1.029	0.3035
		0.8945		**0.3344**		**0.3314**		**0.8945**
4	-1.1503	0.25	-2.3344	0.0459	-1.206	0.388	-1.12	0.2005
	-0.3186	0.25	-0.7420	0.4541	**0.518**	**0.258**	-0.56	0.2995
	0.3186	**0.25**	**0.7420**	**0.4541***	0.911	0.351	**0.56**	**0.2995**
	1.1503	0.25	2.3344	0.0459	5.060	0.003	1.12	0.2005
		0.3778		**0.0408**		**0.1727**		**0.1404**
5	-1.2816	0.20	-2.8570	0.0113	-1.395	0.293	-1.42	0.1434
	-0.5244	0.20	-1.3556	0.2221	0.00	0.0005	-0.71	0.2179
	0.00	0.20	0.00	0.5333	0.403	0.606	0.00	0.2774
	0.5244	**0.20**	**1.3556**	**0.2221**	**1.073**	**0.062***	**0.71**	**0.2179**
	1.2816	0.20	2.8570	0.0113	2.631	0.037	1.42	0.1434
		0.1675		**0.0989**		**0.0081**		**0.0537**
6	-1.3830	0.1667	-3.3243	0.0026	-1.903	0.023	-1.35	0.1303
	-0.6745	0.1667	-1.8892	0.0886	-1.528	0.178	-0.93	0.1195
	-0.2104	0.1667	-0.6167	0.4088	-0.376	0.241	-0.45	0.2502
	0.2104	0.1667	**0.6167**	**0.4088**	0.521	0.486	0.45	0.2502
	0.6745	**0.1667**	1.8892	0.0886	**2.064**	**0.066**	**0.93**	**0.1195***
	1.3830	0.1667	3.3243	0.0026	3.063	0.005	1.35	0.1303
		0.0703		**0.1024**		**0.6077**		**0.0011**

Table 4: Approximation of a $N(0,1)$ distribution under different discretization methods. (Part A)

In the first column are reported the number of mass points used to approximate a normal $N(0,1)$ continuous distribution. In columns 2 to 4 are reported the values and the probabilities of the mass points and the approximation errors $e(Z,\tilde{Z})$ for every considered discretization technique: the standard "bracket mean" method, the Gaussian quadrature of Miller and Rice (1983), the distance minimization (using the first four central moments) proposed by Høyland and Wallace (2001), and the Wasserstein distance minimization (e.g. Pflug (2001), among others). With a star are marked the points among the different methods which minimize the approximation error.

	"Bracket Mean"		Gaussian Quadrature		"Distance Minimization"		Wasserstein-distance	
7	-1.4652	0.1429	-3.7504	0.0005	-1.495	0.237	-1.62	0.0885
	-0.7916	0.1429	-2.3668	0.0308	-1.419	0.003	-1.08	0.1205
	-0.3661	0.1429	-1.1544	0.2401	-0.184	0.142	-0.54	0.1846
	0.00	0.1429	0.00	0.4571	**0.124**	**0.007**	0.00	0.2128
	0.3661	0.1429	**1.1544**	**0.2401**	0.226	0.368	0.54	0.1846
	0.7916	**0.1429**	2.3668	0.0308	1.084	0.221	**1.08**	**0.1205***
	1.4652	0.1429	3.7504	0.0005	2.884	0.022	1.62	0.0885
		0.0245		**0.0246**		**0.6649**		**0.0092**
8	-1.5341	0.125	-4.1445	0.0001	-1.304	0.317	-1.56	0.0861
	-0.8871	0.125	-2.8025	0.0096	-0.839	0.048	-1.17	0.0787
	-0.4888	0.125	-1.6365	0.1172	-0.645	0.00001	-0.78	0.1145
	-0.1573	0.125	-0.5391	0.3730	0.639	0.610	-0.39	0.2207
	0.1573	0.125	0.5391	0.3730	**1.194**	**0.004**	0.39	0.2207
	0.4888	0.125	**1.6365**	**0.1172**	2.014	0.013	0.78	0.1145
	0.8871	**0.125***	2.8025	0.0096	3.000	0.00001	**1.17**	**0.0787**
	1.5341	0.125	4.1445	0.0001	3.802	0.008	1.56	0.0861
		0.0050		**0.2653**		**0.0357**		**0.0288**
9	-1.5932	0.111	-4.5127	0.0001	-2.530	0.034	-1.60	0.0721
	-0.9674	0.111	-3.2054	0.0028	-1.935	0.00001	-1.32	0.0636
	-0.5895	0.111	-2.0768	0.0499	-0.935	0.217	-0.88	0.1189
	-0.2822	0.111	-1.0233	0.2441	-0.914	0.104	-0.44	0.1583
	0.00	0.111	0.00	0.4063	0.088	0.374	0.00	0.1742
	0.2822	0.111	**1.0233**	**0.2441**	**0.985**	**0.181***	0.44	0.1583
	0.5895	**0.111**	2.0768	0.0499	1.909	0.090	0.88	0.1189
	0.9674	0.111	3.2054	0.0028	3.221	0.0001	**1.32**	**0.0636**
	1.5932	0.111	4.5127	0.0001	4.063	0.00002	1.60	0.0721
		0.1198		**0.0023**		**0.0002**		**0.0826**

Table 5: Approximation of a $N(0,1)$ distribution under different discretization methods. (Part B)

	"Bracket Mean"		Gaussian Quadrature		"Distance Minimization"		Wasserstein-distance	
3	0.1823	0.333	0.416	0.711	0.111	0.416	0.58	0.5810
	0.6931	**0.333**	**2.294**	**0.279**	**1.402**	**0.545***	**1.16**	**0.1844**
	1.7918	0.333	6.290	0.010	4.867	0.039	1.74	0.2346
		0.9015		**0.1074**		**0.0869**		**0.2492**
4	0.1335	0.25	0.323	0.603	0.081	0.161	0.52	0.5416
	0.4700	0.25	**1.746**	**0.357***	0.490	0.509	1.04	0.2059
	0.9808	**0.25**	4.537	0.038	**2.024**	**0.311**	**1.56**	**0.1105**
	2.0794	0.25	9.395	0.0005	5.579	0.019	2.08	0.1620
		0.4391		**0.0011**		**0.025**		**0.0291**
5	0.1054	0.20	0.264	0.522	0.034	0.182	0.47	0.5059
	0.3567	0.20	**1.413**	**0.399**	0.302	0.002	0.94	0.1853
	0.6931	0.20	3.596	0.076	0.754	0.662	**1.71**	**0.1158***
	1.2040	**0.20**	7.086	0.004	**3.173**	**0.153**	1.88	0.0724
	2.3026	0.20	12.641	0.0002	9.962	0.0008	2.35	0.1206
		0.2122		**0.0817**		**0.6325**		**0.0034**
6	0.0870	0.1667	0.223	0.4590	0.043	0.006	0.42	0.4674
	0.2877	0.1667	1.189	0.4170	0.389	0.646	0.84	0.1827
	0.5390	0.1667	**2.993**	**0.1134**	**1.617**	**0.214***	1.26	0.1200
	0.8755	0.1667	5.775	0.0104	2.125	0.071	1.68	0.0788
	1.3863	**0.1667**	9.838	0.0003	3.953	0.062	**2.10**	**0.0518**
	2.4849	0.1667	15.983	0.0001	11.09	0.0004	2.52	0.0993
		0.0945		**0.5021**		**0.0162**		**0.043**

Table 6: Approximation of a $exp(1)$ distribution under different discretization methods. (Part A)

In the first column are reported the number of mass points used to approximate a normal $exp(1)$ continuous distribution. In columns 2 to 4 are reported the values and the probabilities of the mass points and the approximation errors $e(Z, \tilde{Z})$ for every considered discretization technique: the standard "bracket mean" method, the Gaussian quadrature of Miller and Rice (1983), the distance minimization (using the first four central moments) proposed by Høyland and Wallace (2001), and the Wasserstein distance minimization (e.g. Pflug (2001), among others). With a star are marked the points among the different methods which minimize the approximation error.

	"Bracket Mean"		Gaussian Quadrature	"Distance Minimization"		Wasserstein-distance	
7	0.0741	0.1429		0.459	0.0005	0.39	0.4429
	0.2412	0.1429		0.513	0.797	0.78	0.1799
	0.4418	0.1429		**2.513**	**0.109**	1.17	0.1218
	0.6931	0.1429		3.337	0.091	1.56	0.0825
	1.0296	0.1429		4.263	0.002	**1.95**	**0.0558***
	1.5404	**0.1429**		6.200	0.00005	2.34	0.0378
	2.6391	0.1429		12.33	0.0003	2.73	0.0793
		0.0344			**0.2074**		**0.0119**
8	0.0645	0.125		0.171	0.456	0.36	0.4173
	0.2076	0.125		**1.325**	**0.009**	0.72	0.1761
	0.3747	0.125		1.406	0.466	1.08	0.1229
	0.5754	0.125		2.326	0.028	1.44	0.0858
	0.8267	0.125		4.586	0.022	**1.80**	**0.0598***
	1.1632	0.125		4.685	0.017	2.16	0.0418
	1.6740	0.125		7.390	0.001	2.52	0.0291
	2.7726	**0.125**		11.895	0.00002	2.88	0.0672
		0.3558			**0.1281**		**0.00003**
9	0.0572	0.111		0.076	0.182	0.34	0.3995
	0.1823	0.111		0.435	0.400	0.68	0.1731
	0.3254	0.111		**1.321**	**0.110**	1.02	0.1232
	0.4925	0.111		1.800	0.267	1.36	0.0877
	0.6931	0.111		3.574	0.006	1.70	0.0624
	0.9445	0.111		4.638	0.033	**2.04**	**0.0444***
	1.2809	**0.111**		6.634	0.0009	2.38	0.0316
	1.7918	0.111		7.196	0.0002	2.72	0.0225
	2.8904	0.111		9.509	0.0001	3.06	0.0556
		0.1559			**0.1305**		**0.0284**

Table 7: Approximation of a $exp(1)$ distribution under different discretization methods. (Part B)

	"Bracket Mean"		Gaussian Quadrature		"Distance Minimization"		Wasserstein-distance	
3	-1.0705	0.333	-3.8730	0.0556	-2.996	0.114	-1.18	0.2904
	0.00	**0.333**	**0.00**	**0.8889**	**0.264**	**0.865***	**0.00**	**0.4192**
	1.0705	0.333	3.8730	0.0556	5.200	0.022	1.18	0.2904
		0.9627		**0.9627**		**0.5135**		**0.9627**
4	-1.3009	0.25			-5.367	0.019	-1.26	0.1940
	-0.3367	0.25			-0.294	0.856	-0.63	0.3060
	0.3367	**0.25**			**1.362**	**0.008***	**0.63**	**0.3060**
	1.3009	0.25			2.948	0.117	1.26	0.1940
		0.4170				**0.0470**		**0.1378**
5	-1.4759	0.20			-3.915	0.054	-1.72	0.1267
	-0.5594	0.20			-0.028	0.126	-0.86	0.2158
	0.00	0.20			-0.008	0.763	0.00	0.3149
	0.5594	**0.20**			**0.868**	**0.0001***	**0.86**	**0.2158**
	1.4759	0.20			3.834	0.057	1.72	0.1267
		0.1897				**0.0269**		**0.0292**
6	-1.6176	0.1667			-3.888	0.055	-1.62	0.1175
	-0.7267	0.1667			-0.924	0.005	-1.08	0.1099
	-0.2217	0.1667			-0.011	0.058	-0.54	0.2726
	0.2217	0.1667			0.003	0.826	0.54	0.2726
	0.7267	**0.1667**			**1.479**	**0.0001**	**1.08**	**0.1099***
	1.6176	0.1667			3.868	0.056	1.62	0.1175
		0.0814				**0.0886**		**0.00005**

Table 8: Approximation of a *student-t*(5) distribution under different discretization methods. (Part A)

In the first column are reported the number of mass points used to approximate a normal *student-t*(5) continuous distribution. In columns 2 to 4 are reported the values and the probabilities of the mass points and the approximation errors $e(Z, \tilde{Z})$ for every considered discretization technique: the standard "bracket mean" method, the Gaussian quadrature of Miller and Rice (1983), the distance minimization (using the first four central moments) proposed by Høyland and Wallace (2001), and the Wasserstein distance minimization (e.g. Pflug (2001), among others). With a star are marked the points among the different methods which minimize the approximation error.

	"Bracket Mean"		Gaussian Quadrature	"Distance Minimization"		Wasserstein-distance	
7	-1.7372	0.1429		-3.956	0.052	-2.04	0.0749
	-0.8610	0.1429		-1.310	0.008	-1.36	0.1023
	-0.3876	0.1429		-0.012	0.105	-0.68	0.1966
	0.00	0.1429		-0.015	0.739	0.00	0.2523
	0.3876	0.1429		-0.033	0.027	0.68	0.1966
	0.8610	**0.1429***		**0.744**	**0.013**	**1.36**	**0.1023**
	1.7372	0.1429		3.837	0.057	2.04	0.0749
		0.0289			**0.0730**		**0.0464**
8	-1.8409	0.125		-3.987	0.051	-1.88	0.0804
	-0.9735	0.125		-0.901	0.0005	-1.41	0.0660
	-0.5204	0.125		-0.293	0.053	-0.94	0.1097
	-0.1655	0.125		-0.043	0.724	-0.47	0.2439
	0.1655	0.125		0.129	0.061	0.47	0.2439
	0.5204	0.125		0.348	0.052	**0.94**	**0.1097**
	0.9735	**0.125***		**0.920**	**0.0001**	1.41	0.0660
	1.8409	0.125		3.793	0.059	1.88	0.0804
		0.0060			**0.0146**		**0.0109**
9	-1.9327	0.111		-4.063	0.048	-2.28	0.0513
	-1.0705	0.111		-1.272	0.0001	-1.71	0.0554
	-0.6313	0.111		-0.546	0.011	-1.14	0.1091
	-0.2979	0.111		-0.183	0.017	-0.57	0.1777
	0.00	0.111		-0.038	0.110	0.00	0.2129
	0.2979	0.111		-0.037	0.648	0.57	0.1777
	0.6313	**0.111**		-0.034	0.101	**1.14**	**0.1091***
	1.0705	0.111		**0.898**	**0.001**	1.71	0.0554
	1.9327	0.111		3.705	0.063	2.28	0.0513
		0.1370			**0.0194**		**0.0029**

Table 9: Approximation of a *student-t*(5) distribution under different discretization methods. (Part B)

4 Definition of the probability set

In section 3, by showing different discretization techniques, we assumed complete information about the probability set and a perfectly known probability density function. Unfortunately, as pointed out in the introduction, by the definition of the probability set we are normally faced with an incomplete information about the measure through which the stochastic program is formulated.

In this section we present the methodology used in the algorithm for the scenario generation, to compute the probability set of underlying asset prices (returns), and the links between the objective probability measure and the risk-neutral measure needed to price options. We could simply assume a given functional form to describe the density function of asset returns, like for example a normal or a student-t distribution, and then discretize it with one of the techniques shown is section 3 in order to obtain the desired number of scenarios. The assumption of a specific density has however, several drawbacks. In particular asset returns distributions are not constant over time and depend on the given market conditions. Moreover, empirical evidences like *'volatility clustering'*, *'fat-tails'* or *'leverage effects'* are unlikely to be captured by the above cited distributions. For these reasons we prefer to infer the probability set directly from observable market data. The goal is to infer satisfactory probability densities such that

- No particular distribution form has to be imposed.
- The well demonstrated existence of volatility clustering is reflected by our distribution.
- In the case of multiple assets, the joint distributions take into account the correlations among those.
- Current (and future possible) market conditions are embedded in the probability densities.

To obtain a distribution satisfying these conditions, we propose to use a similar methodology presented by Barone-Adesi, Giannopoulos, and Vosper (1999). The authors simulate future asset price densities based on the combination of GARCH modeling and historical portfolio returns. This combination of parametric and non-parametric methods, allows us to satisfy the conditions listed above. The GARCH model allows us to be free from imposing a functional form of the distribution, and to describe market volatility clustering. Using historical data allows us to consider asset correlation and current market conditions in the final distributions, using the empirical (historical) distribution of the return series.

GARCH modeling has several advantages in the computation of the risk-neutral probability set needed for option valuation. In contrast to stochastic differential equation models so frequently found in the theoretical financial literature, GARCH models are discrete time stochastic difference equation systems. Since virtually all economic time series data are recorded only at discrete intervals, and a discrete time GARCH likelihood function is usually easy to compute and maximize, empiricists have favored this approach. By contrast, the likelihood of a nonlinear stochastic differential equation system observed at discrete intervals can be very difficult to derive, especially when there are unobservable state variables (Nelson (1990)). Moreover, although continuous time stochastic volatility

models assume that volatility is observable, it is impossible to exactly filter a volatility variable from discrete observations of asset prices in a continuous time stochastic volatility model. Consequently it is not possible to compute out-of-sample options valuation errors from the history of asset returns. Also the unobservability of volatility implies that one has to use implied volatilities computed from option prices to value other options. Holding the model parameters constant through time, this approach requires estimating numerous implied volatilities from option records, one for every date and is computationally very burdensome in a long time series of option records (Heston and Nandi (2000)).

4.1 Barone-Adesi, Giannopoulos, Vosper historical simulation

For an appropriate employment of simulation, the random variables have to be drawn from stationary distributions. Montecarlo simulations assume a particular distributional form, imposing the structure of the returns that they were supposed to investigate. Historical simulations usually sample from past data with equal probability. Therefore they are appropriate only if returns are i.i.d. (independent and identically distributed), an assumption violated by volatilities changing over time. Barone-Adesi et al. show a technique to render the returns i.i.d. and an appropriate way to use them to simulate future probability density functions.

Suppose that asset returns follow a standard GARCH(1,1) process

$$r_t = \mu + \varepsilon_t \tag{19}$$
$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \tag{20}$$

where ω, α, $\beta > 0$, $\alpha + \beta < 1$, μ determines the constant return (continuously compounded), while ε_t is the random residual with $\varepsilon_t = \sqrt{h_t} v_t$, $v_t \sim i.i.d.(0,1)$.

Because of the structure of the volatility h_t, we cannot simulate future asset returns by directly drawing past returns (r) or past random residuals (ε). If however we divide the estimated residual $\hat{\varepsilon}_t$ by the corresponding estimated daily volatility, $\sqrt{\hat{h}_t}$, what we obtain is the series of standardized residual returns

$$e_t = \frac{\hat{\varepsilon}_t}{\sqrt{\hat{h}_t}}$$

which under the GARCH hypothesis are independent and identically distributed (i.i.d.) and therefore suitable for historical simulation. Historical standardized innovations can be drawn randomly (with replacement) and, after being scaled with current volatility, may be used as innovations in the conditional mean (19) and variance (20) equations to generate pathways for future prices and variances respectively (see Barone-Adesi et al. (1999), p. 586). We can so simulate future asset returns from the i.i.d. standardized residuals in the following way:

Step 1: draw T standardized residual returns $e^* = \{e_1^*, e_2^*, \cdots, e_T^*\}$ from a data set Θ with $e_i \in \Theta$, where $i = 1, ..., T$, and T are the days in the future for which we want simulate the probability density function, and compute the next day residual

$$\varepsilon^*_{t+1} = e^*_1\sqrt{h_{t+1}}$$

with h_{t+1} known from the estimation of the GARCH model.

Step 2: compute the next day asset price, from the currently known price

$$p^*_{t+1} = p_t + p_t(\hat{\mu} + \varepsilon^*_{t+1})$$

Step3: finally calculate the future volatilities from equation (20)

$$\sqrt{h^*_{t+i}} = \sqrt{\hat{\omega} + \hat{\alpha}(\varepsilon^*_{t+i-1})^2 + \hat{\beta}h^*_{t+i-1}} \quad i \geq 2.$$

where, for $i = 2$, $h^*_{t+i-1} = h_{t+1}$ is directly known from the estimation of the GARCH process.

Now repeat the points (1) to (3), computing ε^*_{t+i}, p^*_{t+i} and $\sqrt{h^*_{t+i}}$ for every period till period T. This procedure can be replicated a large number of times, obtaining for each time horizon i the empirical density function of the asset prices.

The advantage of this methodology is that in case of multiple assets their correlation is taken implicitly into account without restricting their values over time or the need to compute the correlation matrix explicitly. In case of multiple assets the authors draw for each underlying asset $j = 1, ..., N$ in the portfolio, the T standardized residual returns

$$\begin{aligned}
\text{Asset 1:} \quad & e^*_1 = \{e_1, e_2, \ldots, e_T\}_1 \\
\text{Asset 2:} \quad & e^*_2 = \{e_1, e_2, \ldots, e_T\}_2 \\
& \vdots \\
\text{Asset N:} \quad & e^*_N = \{e_1, e_2, \ldots, e_T\}_N
\end{aligned}$$

At every draw, all the respective standardized residual returns for each of the N different assets are seized, so that comovements between assets are considered. In other words we draw T dates from the past and for every date the N standardized residual. We use them to compute the future asset prices and future volatilities for every asset, as done in point (2) and (3). We then draw other T dates form the history of standardized residual and compute again the future prices and volatilities. This procedure is repeated a large number of times (in our case we will simulate ten thousand prices for every period) so that asset prices densities can be inferred.

The main insight of the Barone-Adesi et al. (1999) method is that it is possible to capture conditional heteroskedasticity (volatility clustering) in the data and still be somewhat unrestrictive about the shape of the distribution of the factor returns. Thus the method appears to combine the best elements of conditional volatility models with the best elements of the historical simulation method (Pritsker (2001)).

Pritsker, however, also pointed out three main critiques about this method. First, the assumption that volatility depends only on a risk factor's own past lags, and its

own past lagged volatility can be unrealistic depending on whether there is a single risk factor, or many. For instance, it is reasonable to think that assets belonging to a common market segment show some dependence structure and that one asset can be influenced and predicted by past values of some others. Second, the assumption that ε_t is i.i.d. implies that the correlation of the risk factors is fixed through time. This assumption is also likely to be violated in practice. Finally, Pritsker found in his investigation that the Barone-Adesi et al. method fails to accurately compute the tails of the distribution for long time horizons. This is due to a lack of extreme outliers in the filtered data set. A possible and simple way to avoid this last problem could be the use of a longer span of historical data, which may increase the number of outliers. Unfortunately, this will increase the problem of fixed conditional correlation shown in point two.

Audrino and Barone-Adesi (2005) present a procedure based on functional gradient descent (FGD) estimation for the volatility matrix (Audrino and Bühlmann (2003)) which should contrast well the critiques made about the Barone-Adesi et al. method. First, the FGD technique allows for the use of cross-terms as predictor variables, which solves the first problem. Second, the authors do not assume fixed conditional correlations, but only constant conditional correlations in a rolling (i.e. not fixed) time-window of about three years of data, using to model the dynamics of the multivariate return series the constant conditional correlation (CCC) model firstly proposed by Bollerslev (1990). This solves partially the second problem. Finally, they found through simulation that their model describes more accurately the tails of the distribution.

Because of its simplicity and its intuitiveness we decided, however, to apply directly the Barone-Adesi et al. (1999) method. In our empirical analysis we compute scenarios solely for one underlying asset and several options on this asset, thus avoiding the problem of cross-dependency of multiple risk factors. Moreover, the main focus of our work is the computation of contingent claims scenarios. We leave therefore the implementation and the analysis of the Audrino and Barone-Adesi (2005) procedure in our scenario generation algorithm for future research.

The Barone-Adesi, Giannopoulos, and Vosper (1999) historical simulation is very flexible, and can be used with any GARCH specification. In order to obtain the underlying asset prices probability densities, we analyze three different GARCH processes used in the past literature to describe asset returns. The first two are similar to the nonlinear asymmetric GARCH (NGARCH) and VGARCH processes studied by Engle and Ng (1993) and Duan (1995). These are used by Ritchken and Travor (1999) (from now on also called the RT model) and Heston and Nandi (2000) (HN model). The third is the threshold GARCH of Glosten, Jagannathan, and Runkle (1993) (GJR model). Unlike the standard GARCH(1,1) presented above, these three processes have an asymmetry component, which allows reflecting empirically observed "leverage effect", i.e. the negative correlation between return and volatility innovation. Moreover the RT and HN models have the advantage that under suitable preference restrictions, a local risk-neutralized probability measure can be analytically established under which option prices can be computed as simple discounted expected values.

4.2 The RT and HN GARCH processes

Let the asset price at time t be denoted by S_t, and the return by the difference of the logarithm of the price at time t and time $t-1$, i.e. $r_t = log(S_t/S_{t-1})$. The RT GARCH is

$$\begin{aligned} r_t &= r_f + \lambda\sqrt{h_t} - \frac{1}{2}h_t + \sqrt{h_t}v_t \\ h_t &= \omega + \beta h_{t-1} + \alpha h_{t-1}(v_{t-1} - \gamma)^2, \end{aligned} \tag{21}$$

while the first-order HN process is

$$\begin{aligned} r_t &= r_f + \lambda h_t + \sqrt{h_t}v_t \\ h_t &= \omega + \beta h_{t-1} + \alpha(v_{t-1} - \gamma\sqrt{h_{t-1}})^2, \end{aligned} \tag{22}$$

where r_f is the continuously compounded riskless rate of return over the period from $t-1$ and t, v_t is a standard normal disturbance and $h_t|I_{t-1}$ is the conditional variance of the return between $t-1$ and t, known from the information set at time $t-1$. The unit risk premium for the asset is λ. The expected spot return exceeds the riskless rate by an amount proportional to the variance h_t. To ensure that the volatility stays positive ω, α and β should be nonnegative. The γ parameter results in asymmetric influence of shocks (the so called "leverage effect").

A feature of these two GARCH processes is that they have an interesting continuous time limit. For both models the variance process h_t converges weakly to a variance process, $\nu(t)$ which is the square-root process of Feller (1951), Cox, Ingersoll, and Ross (1985) and Heston (1993)[2]

$$d\nu = \kappa(\theta - \nu)dt + \sigma\sqrt{\nu}dz$$

as the time interval between $t-1$ and t shrinks to zero, where $z(t)$ is a Wiener process. Consequently the two RT and HN processes contain Heston (1993) continuous-time stochastic volatility model as a special case.

The processes in equation (21) and (22) can be used as previously seen, to simulate the next T periods densities. This is done for all underlying assets in the portfolio under the real or objective probability measure $\mathcal{P}$. We could also use the same processes to infer the contingent densities under the risk-neutral measure $\mathcal{Q}$. In this case however, either the parameters are recalibrated in order to be consistent with empirically observed contingent market prices (a similar technique is used by Barone-Adesi, Engle, and Mancini (2004)), or a consistent discount factor r (different from the risk-free rate) has to be found.

A third possibility is given by the locally risk-neutral valuation relationship (LRNVR). Following Duan (1995), a pricing measure $\mathcal{Q}$ is said to satisfy the locally risk-neutral valuation relationship if measure $\mathcal{Q}$ is mutually absolutely continuous with respect to measure $\mathcal{P}$, $S_t/S_{t-1}|I_{t-1}$ distributed lognormally under $\mathcal{Q}$,

$$E^{\mathcal{Q}}(S_t/S_{t-1}|I_{t-1}) = e^r$$

[2]See Heston and Nandi (2000), appendix B for details.

and

$$Var^{\mathcal{Q}}(log(S_t/S_{t-1})|I_{t-1}) = Var^{\mathcal{P}}(log(S_t/S_{t-1})|I_{t-1})$$

almost surely with respect to measure $\mathcal{P}$. This definition of the LRNVR, requires that the conditional variances under the two measures are equal. This is derivable because one can observe and hence estimate the conditional variance under $\mathcal{P}$. If we assume that the LRNVR holds, and that the one-period ahead conditional variance is invariant with respect to a change to the risk-neutralized pricing measure[3], we can rewrite the two processes of equation (21) and (22), under the local risk-neutralized measure as

$$\begin{aligned} r_t &= r_f - \frac{1}{2}h_t + \sqrt{h_t}\xi_t \\ h_t &= \omega + \beta h_{t-1} + \alpha h_{t-1}(\xi_{t-1} - (\gamma + \lambda))^2, \end{aligned} \tag{23}$$

$$\begin{aligned} r_t &= r_f - \frac{1}{2}h_t + \sqrt{h_t}\xi_t \\ h_t &= \omega + \beta h_{t-1} + \alpha(\xi_{t-1} - (\gamma + \lambda + \frac{1}{2})\sqrt{h_{t-1}})^2, \end{aligned} \tag{24}$$

where ξ_t is a standard normal random variable with respect to the risk-neutralized probability measure[4]. We can now use these two processes to simulate the densities under the measure $\mathcal{Q}$, and use the continuously compounded riskfree rate r_f, to discount future contingent prices.

4.3 The GJR threshold GARCH process

The last GARCH process we examine is the threshold GARCH process of Glosten, Jagannathan, and Runkle (1993). Let the asset price at time t be S_t, and the return by the difference of the logarithm of the price at time t and time $t-1$, $r_t = log(S_t/S_{t-1})$. The GJR GARCH is

$$\begin{aligned} r_t &= \mu + \varepsilon_t \\ h_t &= \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1} + \gamma\Psi_{t-1}\varepsilon_{t-1}^2, \end{aligned} \tag{25}$$

where $\omega, \alpha, \beta > 0$, $\alpha + \beta + \gamma/2 < 1$, μ determines the constant return (continuously compounded) of the underlying asset S_t, $\varepsilon_t = \sqrt{h_t}v_t$, $v_t \sim i.i.d.(0,1)$ and $\Psi_{t-1} = 1$, when $\varepsilon_{t-1} < 0$ and $\Psi_{t-1} = 0$, otherwise. As for the previous two models, γ accounts for the "leverage effect", the stronger impact of "bad news" ($\varepsilon_{t-1} < 0$) rather than "good news" ($\varepsilon_{t-1} \geq 0$) on the conditional variance h_t.

The advantage of this model is the facility to estimate the process parameters. Unfortunately, in this case we cannot derive the threshold GARCH under the local risk-neutralized

[3]It is assumed that the one-period ahead conditional variance is invariant with respect to a change in measure. The unconditional variance or any conditional variance beyond one period is not invariant to the change in measures caused by risk neutralization (Duan (1995)).

[4]See Duan (1995) and Heston and Nandi (2000) for the derivation of the two risk-neutralized processes.

measure as done for the RT and HN models. We therefore use the same process for both the objective probability measure $\mathcal{P}$ and the risk-neutral measure $\mathcal{Q}$. The problem is to find a consistent discount factor needed to obtain contingent claim prices. To bypass this problem, we derive a consistent discount rate from the put-call parity formula. Using the methodology of Shimko (1993), it is possible to estimate the risk-free rate and the dividend yield exploiting the put-call parity equation:

$$c_t - p_t = S_t e^{-q(T-t)} - X e^{-r(T-t)} \tag{26}$$

where c is the price of a European call and p is the price of a European put at time t with the same time-to-maturity $(T-t)$ and the same strike price (X). S_t is the time t price of the underlying security, $e^{-q(T-t)}$ is the dividend discount factor, while $e^{-r(T-t)}$ is the risk-free discount factor. Since the two implicit discount factors arise from existing contingent claims, we may consistently discount option prices recovered using the same GARCH process used to infer underlying prices under the probability measure $\mathcal{P}$.

5 Option prices scenarios

What we have seen so far, especially in sections 3 and 4, are the theoretical aspects needed to define the random variables probability sets under both the objective and the risk-neutral measures, and the techniques to discretize the derived probability densities. These discretization techniques can be used to obtain a limited number of scenarios, needed to implement a stochastic programming model.

While for the underlying assets the whole scenario proceeding is in some way well defined, for contingent claims several aspects need a further analysis. Theoretically, we could implement the same proceeding used for the underlying assets also for the contingent claims: we could simulate, using the risk neutralized GARCH processes, the densities for the underlying assets under the risk-neutral measure and compute the contingent prices at expiration. These densities can then be discretized obtaining a limited number of contingent prices scenarios. This procedure is, however, only partially satisfactory and several problems have still to be solved. First, the procedure explained above allows to compute contingent prices only at expiration when prices depend exclusively on the value of the underlying asset and a given strike level. The price at state i of a call or a put option at expiration is, for example, given by the formulas $(S_i - K)^+$ and $(K - S_i)^+$, respectively. For all other investment periods (differently than for the underlying assets) the contingent prices remain unknown. Second, several no-arbitrage conditions have to be satisfied. In particular in order to avoid any arbitrage possibility the discounted prices of the underlying and the contingent assets have to be equal to the observed market prices. This is not guaranteed, without further specifications, by the procedure used for the underlying assets.

In this section we first introduce two different models, which describe how to infer option price scenarios. The first is proposed by Rubinstein (1994) and Jackwerth and Rubinstein (1996). Assuming a recombining binomial tree for the underlying asset the authors compute the risk-neutral probabilities in such a way that these are consistent with the observed option market prices. The assumption of a binomial tree is crucial and can be unfavourable in case of a large amount of scenarios. The second model is proposed by Schyns, Crama, and Hübner (2003) and has the advantage not to be bound to a strict tree structure like the binomial one. The problem is that we can compute contingent prices only at expiration and not for the entire scenario tree, which should describe the process to go from the option market prices observed today to the option prices at expiration.

The second part of the section is used to explain an alternative methodology which tries to reconcile the advantages of both models in a single framework. In particular, working with a defined tree structure like in the Rubinstein model, allows us to define the entire tree option prices and probabilities (and not just the state contingent prices at expiration). Moreover, the methodology of Schyns et al. allows us to avoid using a full recombining binomial tree, and to work with a more flexible tree structure like a nonrecombining multinomial one.

5.1 The (Jackwerth-) Rubinstein implied binomial tree model

The model presented by Rubinstein (1994) and refined by Jackwerth and Rubinstein (1996) infers risk-neutral probabilities (or state-contingent prices) from the simultaneously observed prices of European options, and uses these probabilities to compute a unique and fully specified recombining binomial tree that is consistent with these probabilities (and, hence, consistent with all the observed option prices). The idea consists in establishing an estimate of the risk-neutral probabilities P'_j with $j = 1, .., J$ and J the total number of scenarios at option expiration, and then solving the following quadratic program in order to find the implied posterior risk-neutral probabilities P_j

$$\min_{P_j} \sum_j (P_j - P'_j)^2 \quad \text{subject to:}$$

$$\sum_j P_j = 1 \quad \text{and} \quad P_j \geq 0 \quad \text{for } j = 0, ..., n$$

$$S^b \geq S \geq S^a \quad \text{where} \quad S = \left(\frac{\delta}{r}\right)^n \sum_j P_j S_j$$

$$C_i^b \geq C_i \geq C_i^a \quad \text{where} \quad C_i = \frac{\sum_j P_j max[0, S_j - K_i]}{r^n} \quad \text{for } i = 1, ..., m.$$

Where S^b (S^a) are the current bid (ask) price of the underlying asset, and C_i^b (C_i^a) the current bid (ask) price for a call i maturing after n periods. In order to setup the model, we need to sample future asset values S_j and the prior risk-neutral probabilities P'_j. For this purpose, Rubinstein suggests to construct a n-step standard binomial tree using the average of the Black-Scholes implied volatilities of the two nearest-the-money call options, and use the asset values and probabilities computed at the final leaves of the tree as prior. Finally r and δ represent, respectively, the riskless interest rate and underlying asset payout return over each binomial period. According to the authors, the obtained P_j are the risk-neutral probabilities, which are, in the least square sense, closest to lognormal that cause the present values of the underlying asset and all the options calculated with these probabilities to fall between their respective bid and ask prices. The model seems quite robust. The authors show that by changing the objective function or prior distribution (using a uniform distribution instead of a lognormal resulting from the binomial lattice) all the cumulative distribution functions converge, assigning about the same cumulative probabilities to values near-the-money.

This method of computing risk-neutral probabilities presents, however, certain limitations. The first and maybe most important, is the limited number of scenarios we can construct, cause the binding assumption of a recombining binomial tree. Replicating the model, we were able to compute around 300 scenarios, before the optimization model broke down. This seems quite a large number of possible scenarios, but it is not, depending on the time window we consider. Suppose for example that we want to build a scenario tree for the price of options which mature one month from now. Assume also that we have in mind a four weeks multinomial scenario tree with five outcomes starting from every node each week. Five outcomes every week for four weeks bring to a total number of final leaves equal to 5^4=625 scenarios. To obtain the same number of scenarios with the Rubinstein

model, we should construct a 624 period binomial tree, which we were not able to solve. Obviously this problem increases with the number of outcomes as with the increase of the option time-to-maturity and the steps we want before reaching the final leaves. A second limitation, presented in Schyns, Crama, and Hübner (2003), is the possibility that the optimization is infeasible. The Rubinstein model assumes that there is no arbitrage opportunity for the scenarios representing the future market and the set of calls under consideration. Schyns et al. argue, however, that this assumption does not necessarily hold, and show an empirical example in which the optimization model reaches no optimal solution.

Suppose now that besides the previously presented model limitations, we have found a satisfying discretized risk-neutral probability distribution of the underlying returns at the time of option expiration. Under some given assumption, Rubinstein shows how to find, working backward from the final risk-neutral probabilities, a unique binomial lattice. We are thus able to find all the up and downward moves and the risk-neutral probabilities for each step of the binomial tree. Once we have the entire structure of the binomial lattice with the probabilities and the up and down moves, it is possible to price every option on the starting underlying S. In fact this model is a generalization of the standard binomial option pricing model in which the up and down moves are constant for the whole tree.

Once the risk-neutral distribution is established, what is left is the computation of the consensus probabilities. It is well known that the subjective probability distribution of ending returns cannot be inferred only from knowledge of its risk-neutral distribution. In the presence of a complete market with a representative investor who maximizes his expected utility, it is however possible to find the ending consensus subjective nodal probabilities. Knowing the ending risk-neutral probabilities P_j, the returns $R_j = S_j/S$ and hypothesising a utility function $U(\delta^n R_j)$ it is possible to find the consensus ending probabilities Q_j by solving the Lagrangian problem:

$$max \sum_j Q_j U(\delta^n R_j) - \lambda \left[\sum_j \left(\frac{P_j}{r^n} \right) \delta^n R_j - 1 \right]$$

where $Q_j > 0$ is the subjective probability the investor attends to state j. The P_j/r^n are often called "state contingent prices". From the first order conditions we obtain

$$Q_j = \frac{\lambda \left(P_j / r^n \right)}{U'(\delta^n R_j)}$$

with the Lagrange multiplier equal to

$$\lambda = \left[\sum \frac{P_j / r^n}{U'(\delta^n R_j)} \right]^{-1}.$$

Having already computed the risk-neutral probabilities P_j, and the future asset returns R_j we have every element to recover the consensus subjective nodal probabilities Q_j. The same technique to go from risk-neutral to consensus probabilities is adopted by Schyns, Crama, and Hübner (2003). Since, however, the Rubinstein model does not always guarantee a solution, they use a different method to infer the ending risk-neutral probabilities and to price options.

5.2 The Schyns, Crama, Hübner (2003) model

Breeden and Litzenberger (1978) show that if options had striking prices infinitely dense on the positive real line, then there would be a perfect relationship between option prices and the probability density function f of the underlying security returns S. This relation is given by the second derivative of the option price C on the strike price K. Suppose the price of a European call option is a continuous function $C(K)$ of the strike price K, then from the option pricing formula (see Cox, Ross, and Rubinstein (1979)) we can derive:

$$C(K) = e^{-r(T-t)} \int_K^\infty (S-K) f(S) dS$$
$$\frac{\partial C(K)}{\partial K} = -e^{-r(T-t)}(1 - F(K))$$
$$\frac{\partial^2 C(K)}{\partial K^2} = e^{-r(T-t)} f(K).$$

The problem is that the observed option prices are only available at discretly spaced strike price levels, with the lowest price well above zero and the highest well below infinity. To solve this problem, Shimko (1993) proposes a method to obtain an analytical expression of $C(K)$ taking the smile effect into account. The author computes first the implied volatilities across strike prices of options over a given underlying asset. Second, he smoothes the implied volatilities $\sigma(K)$ according to a quadratic relation

$$\sigma(K) = \beta_0 + \beta_1 K + \beta_2 K^2$$

and computes the coefficient by a least square method, obtaining an analytical expression of the implied volatility as a function of the strike prices K.

This smoothed volatility can be used to find the smoothed call price (using the Black-Scholes formula).The smoothed call prices can then be differentiated in order to find the values of the density function and the cumulative distribution function for each possible value of K. In other words we insert the implied volatilities in the Black-Scholes option pricing formula

$$C(K) = Se^{-\delta(T-t)} N(d_1) - Ke^{-r(T-t)} N(d_2)$$

where

$$d_1 = \frac{ln(\frac{S}{K}) + (r - \delta + \sigma^2(K)/2)(T-t)}{\sigma(K)\sqrt{T-t}}$$
$$d_2 = d_1 - \sigma(K)\sqrt{T-t},$$

with $N(.)$ the normal cumulative distribution function. The risk-neutral probability density function $f(K)$ is obtained computing the second derivative of $C(K)$ with respect to K.

The continuous risk-neutral probability density function can be discretized to obtain the desired number of scenarios. Schyns, Crama, and Hübner (2003) propose a "bracket mean" method[5] (called in the paper "stratified sampling" or "stylized sampling"), reducing the density into N equally probable future scenarios of the underlying asset S. These scenarios are used as prior risk-neutral probabilities in a quadratic minimization problem. The idea consists in computing the today option prices, which are as close as possible to the appropriate target values (in concrete the observed market option prices) finding the posterior risk-neutral probabilities, consistent with general no-arbitrage conditions. The model is:

$$\begin{aligned} &\min \sum_{i=1}^{I} \left(\frac{\tilde{C}_i - Target_i}{Target_i} \right)^2 \\ \text{s.t.} \quad &\begin{pmatrix} 1 \\ S_0 \\ \tilde{C}_1 \\ \vdots \\ \tilde{C}_I \end{pmatrix} = \begin{pmatrix} e^{rn} & \cdots & e^{rn} \\ S_1 e^{\delta n} & \cdots & S_N e^{\delta n} \\ C_{1,1} & \cdots & C_{1,J} \\ \vdots & \vdots & \vdots \\ C_{I,1} & \cdots & C_{I,J} \end{pmatrix} \begin{pmatrix} \tilde{\pi}_1 \\ \tilde{\pi}_2 \\ \tilde{\pi}_3 \\ \vdots \\ \tilde{\pi}_J \end{pmatrix} \\ &\tilde{C}_i \geq 0 \\ &\tilde{\pi}_j \geq 0 \end{aligned} \tag{27}$$

where $j = 1, ..., J$ and $i = 1, ..., I$ are respectively the different scenarios and the different options indices. δ and r are the dividend yield and the risk-free return, while $e^{rn}\pi_j$ and n are the posterior risk neutral probabilities and the periods to maturity, respectively. The final option prices are not random, but instead are known variables. This is because it is assumed that the problem is solved over the period from today to the time-to-maturity, n days from now. We can, therefore, easily compute the call (put) option prices for each final leaf just computing the $max\{S - K, 0\}$, (or $max\{K - S, 0\}$) for each given strike price. Would the final leaves not be at expiration, we would have the problem of how to calculate the final option prices, needed to solve the optimization problem. While for the Rubinstein implied binomial tree model, every node of the tree is characterized by its price and probability, here no structure for the scenario tree is defined, and only the starting prices (at time 0) and the final leaves option expiration payouts (after n periods) are computed. To infer the intermediate nodes we should first define a structure for the scenario tree and second, solve for every stage the optimisation (27), with the problem of how define the option price target values ($Target_i$) for the intermediate nodes of the tree.

Once obtained the risk-neutral probabilities, the authors use the same technique used by Rubinstein to compute the consensus probabilities.

5.3 An alternative scenario generation model

The two presented models start from information about the option market to infer the scenarios under both probability measures. The authors use the option market prices and their implied volatilities as unique input, completely ignoring information rising from the

[5] See section 3.2

stock market (beside the today asset price). They limit their observations to a restricted number of observable option prices and justify their choice adducing that the information derived from current options has a "forward-looking" nature, since it considers today prices with a future time-to-maturity.

What we try to do with our model is to embed information rising from the stock market in the input data set, arguing that this should increase the quality of the computed scenarios. As seen in the previous section, hypothesizing a GARCH process for the underlying assets we were able to infer for every desired period in the future the probability densities under both, the objective and the risk-neutral probability measure. Discretizing these densities we can describe the entire underlying asset evolution with a flexible multinomial tree[6] for the periods until the option time-to-maturity. This has the advantage that to infer future asset values and prior risk-neutral probabilities, we are not constrained to use a recombining binomial tree, like in the Rubinstein model, but instead an unrestricted (at least theoretically) non-recombining multinomial tree. For this resulting multinomial scenario for every node we know the underlying prices and the occurrence probability (i.e. the probability to pass from the initial node to each other defined node in the tree).

With the obtained underlying prices we can derive for every final leaf the option values at expiration and the related occurrence probabilities. Having computed the entire underlying scenario tree and under the hypothesis that the structure of the tree is maintained for the option market, we can use the option values at expiration to recover the option prices at every node of the tree. Using a similar minimization problem as the one presented by Schyns et al. we compute the risk neutral probabilities so that the today option prices are as close as possible to the observed market prices, and no arbitrage condition is allowed. Knowing the entire structure of the multinomial scenario tree, the final leaves option prices and the risk-neutral probabilities, it is then possible to recover option prices and occurrence probabilities for every node of the tree.

This new methodology eliminates several important limitations of the two previous models. First, the structure of the tree has not to be strictly imposed as for the recombining binomial tree in the Rubinstein model nor the intermediate nodes remain indeterminate as in the Schyms, Crama Hübner model. In fact the number of periods and nodes is potentially completely unrestricted (up to a tolerable computational time limit) and can be completely determined, without the necessity of new inputs or additional restrictions. Second, we do not need to impose a utility function to pass from the risk-neutral probabilities to the consensus one. Once the scenarios and the structure of the tree are determined for the "real world", we can define the entire option scenario tree simply by solving a standard minimization problem. Finally, beside the information raised from the option market prices, we also consider information from the underlying price process. We therefore do not limit our inputs to few option prices, but are able to compute scenarios which indirectly consider the volatility clustering of the stocks, the current market conditions and in case of multiple assets, the correlations among these.

[6] In part II we present the algorithm and each step to construct such a scenario tree in detail.

Part II
Algorithm and empirical results

6 The scenario generation Algorithm

In the first part of the work we gave a general overview of the existing literature concerned with the generation of asset scenarios, and to identify the problems we are faced with in such an exercise.

In this second part we describe in detail the algorithm used to construct multinomial scenario trees needed as input to a multiperiod stochastic programming model. In the next section we then present several interesting empirical results. The framework is similar to the one of the financial planning example of section 2.2.1. We analyze a portfolio consisting of one or more underlying assets and several options on these assets. In order to optimize the composition of our portfolio with a stochastic program, we have to generate reasonable scenarios describing the possible future prices of all considered assets and the probability under which a determinate price may occur.

In a similar way as we have structured the theoretical part (from section three to five), the overall proceedings can be synthesized in the following 3 main steps:

- Identify the process describing the future returns of the underlying assets, and compute the probability density functions of these returns.
- Discretize the densities to obtain a limited number of scenarios, and construct a multinomial scenario tree for the underlying assets, defining the prices (returns) and the respective probabilities for each node of the tree.
- Compute the multinomial scenario trees for the contingent claims, so that the structure of the underlying tree is maintained, the no-arbitrage conditions are satisfied, at expiration option prices are consistent with future expected underlying prices and option strikes, and finally, the theoretical option prices are as close as possible to the empirical prices observed in the market.

6.1 Underlying scenario tree

To understand the procedure, consider the following simplified market conditions (further generalizations can be easily included and are discussed below). Let the investment opportunity set consist only of a unique underlying asset and of different options on this asset, with the same time to maturity, T days from now. As for the financial planning example, assume we want to change the weights of our portfolio every ν days, so that we have $D = T/\nu$ investment periods.

The goal is to construct a D-stage multinomial scenario tree for the underlying asset and for the related options, so that (1) at every future investment period $d = 1, \ldots, D$ we obtain *'good'* scenarios which describe well the uncertainty surrounding market evolutions, and (2) over different periods, the connected nodes are related in a *consistent intertemporal relation*. As explained in section 3, the necessity of working with a limited number of scenarios is due to the technical and computational complexity of stochastic programming models. In order to implement and solve such a model in reasonable time we have to restrict the set of possible random events. Remember that in a stochastic program (as opposed to a static program) random variables assume multidimensional data trajectories in a probability space. Letting the dimension of these trajectories be very high makes

the optimization problem unsolvable or the required computational time unacceptable for practical purposes.

6.1.1 Simulation of the probability densities

Step 1 of the algorithm consists in the identification of the underlying price process, the estimation of all process parameters and the simulation of future price densities. As reported in section 4, we analyze three different GARCH specifications: the GARCH used by Ritchken and Trevor in 1999 (RT-GARCH), the one proposed by Heston and Nandi in 2000 (HN-GARCH) and the threshold GARCH of Glosten, Jagannathan and Runkle (1993) (GJR-GARCH). We use these three GARCH processes, first to emphasize the flexibility of the algorithm, which could theoretically be computed independently of the process specification. Second, we are interested in identifying which model computes the best prior risk-neutral probabilities used to compute posterior option prices.

Let S_t be the closing underlying price at day t and r_t the daily log-return, $r_t := ln(S_t/S_{t-1})$, we can write the three processes as follows.

Ritchken and Trevor GARCH:

$$\begin{aligned} r_t &= r_f + \lambda\sqrt{h_t} - \frac{1}{2}h_t + \sqrt{h_t}v_t \\ h_t &= \omega + \beta h_{t-1} + \alpha h_{t-1}(v_{t-1} - \gamma)^2, \end{aligned} \tag{28}$$

Heston and Nandi GARCH:

$$\begin{aligned} r_t &= r_f + \lambda h_t + \sqrt{h_t}v_t \\ h_t &= \omega + \beta h_{t-1} + \alpha(v_{t-1} - \gamma\sqrt{h_{t-1}})^2, \end{aligned} \tag{29}$$

Glosten, Jagannathan and Runkle GARCH:

$$\begin{aligned} r_t &= r_f + \varepsilon_t \\ h_t &= \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1} + \gamma I_{t-1}\varepsilon_{t-1}^2. \end{aligned} \tag{30}$$

where $\varepsilon_t = \sqrt{h_t}v_t$ and $v_t \sim i.i.d.(0,1)$ is an identical and independently distributed random variable. r_f is the continuously compounded riskless rate of return over the period from $t-1$ and t, and h_t is the conditional variance of the return between $t-1$ and t, known from the information set at time $t-1$.

The first task is the estimation of the parameters of the three processes. We do this with the maximum likelihood estimation (MLE). Once these processes are estimated we use the methodology described by Barone-Adesi, Giannopoulos, and Vosper (1999) and compute the time series of standardized residuals $e_t = \hat{\varepsilon}_t/\sqrt{\hat{h}_t}$ (or equivalently $v_t = \hat{v}_t$). From the data set Θ consisting in the last three years of data (750 observations), we draw ν standardized residuals $e^* = \{e_1^*, e_2^*, \ldots, e_\nu^*\}$ with $e_\tau \in \Theta$, and $\tau = 1, \ldots, \nu$. Under

the GARCH hypothesis these standardized residuals are i.i.d. and therefore suitable for historical simulation.

For each process, we compute then the next ν days residual

$$\varepsilon^*_{t+\tau} = e^*_\tau \sqrt{h_{t+\tau}} \quad \text{with } \tau = 1, \ldots, \nu;$$

the next ν days asset prices (using the asset return equations reported above) and the future volatilities, for each of the three h_t equations.

The advantages (and critiques) of using this methodology are reported in section 4. In particular, in the case of multiple underlying assets the methodology allows us to consider the correlation among the assets without having to estimate the variance-covariance-matrix. Repeating the procedure a sufficient number of times (we simulate ten thousand prices for each period) we can infer the probability density of the underlying asset prices for the first investment date (d=1). Simultaneously, we compute the density of the underlying prices under the risk-neutralized probability measure for the RT and the HN models, using the two risk-neutralized GARCH processes discussed in section 4:

$$\begin{aligned} r_t &= r_f - \frac{1}{2} h_t + \sqrt{h_t}\xi_t \\ h_t &= \omega + \beta h_{t-1} + \alpha h_{t-1}(\xi_{t-1} - (\gamma + \lambda))^2 \end{aligned} \tag{31}$$

$$\begin{aligned} r_t &= r_f - \frac{1}{2} h_t + \sqrt{h_t}\xi_t \\ h_t &= \omega + \beta h_{t-1} + \alpha(\xi_{t-1} - (\gamma + \lambda + \frac{1}{2})\sqrt{h_{t-1}})^2 \end{aligned} \tag{32}$$

for the Ritchken and Trevor and the Heston and Nandi processes, respectively. The procedure to compute the risk neutralized probability densities is exactly the same as the one used to infer the densities under the objective probability measure.

Once we have the probability densities for period d=1 under both probability measures (at least for the RT and HN processes), we have to discretize them in order to obtain the desired number of scenarios for the first investment period. This is a crucial task, and has to be done very carefully, since the quality of the scenarios depends extremely on the discretization procedure.

6.1.2 Discretization procedure

To obtain the desired number of scenarios, at every node of the scenario tree, step 2 of the algorithm consists in the discretization of the probability density function (*pdf*) of the underlying asset prices. A natural criteria to define *'good'* scenarios seems therefore to be the minimization of the approximation error $\mathrm{e}(Z, \tilde{Z})$ defined in section 3:

$$e(Z, \tilde{Z}) := Z(argmin_x \tilde{Z}(x)) - Z(argmin_x Z(x)) = \int_\Omega f(\omega, \tilde{x}^*) P(d\omega) - \int_\Omega f(\omega, x^*) P(d\omega)$$

where $Z(.)$ and $\tilde{Z}(.)$ are the objective functions of the stochastic program under the continuous and the approximated probability measure, respectively. Ω is the set of possible

random events ω, while x^* and $\tilde{x}^*$ are the optimal solutions under the two probability measures. Minimizing the approximation error $e(Z, \tilde{Z})$ we obtain an optimal solution $\tilde{x}^*$ which is as close as possible to the unrestricted one (x^*) nevertheless working with a restricted probability set.

As seen in section 3, because of the difficulty of calculating $e(Z, \tilde{Z})$, we use an upper bound for it (see lemma 1 on page 20 and Pflug (2001))

$$e(Z, \tilde{Z}) \leq 2 \sup_x \left| Z(x) - \tilde{Z}(x) \right| = 2 \sup_x \left| \int_\Omega f(\omega, x) P(d\omega) - \int_\Omega f(\omega, x) \tilde{P}(d\omega) \right|.$$

If the cost functions of the stochastic program $\omega \mapsto f(x, \omega)$ are uniformly Lipschitz of order one, i.e. for all $x \in X$,

$$L_1(f) = \inf \{ L : |f(x, u) - f(x, v)| \leq L |u - v| \} \leq \bar{L}_1 \quad \text{for all } u, v \in \Omega$$

then the problem of minimizing $\sup_x \left| Z(x) - \tilde{Z}(x) \right|$ can be approximated by the problem of minimizing the Wasserstein distance d_1

$$\sup_x \left| Z(x) - \tilde{Z}(x) \right| \leq \bar{L}_1 d_1(P, \tilde{P})$$

with $d_1(F, \tilde{F})$ equal to [7]

$$d_1(P, \tilde{P}) = \sup \left\{ \int f(\omega) dP(\omega) - \int f(\omega) d\tilde{P}(\omega) : L_1(f) \leq 1 \right\}.$$

Theorem 1 on page 28 tells us that given the mass points $z_1, \ldots, z_k$ the Wasserstein distance between the continuous an the approximated distribution can be written as

$$d_1(P, \tilde{P}) = \sum_{i=1}^{k} \int_{\frac{z_{i-1}+z_i}{2}}^{\frac{z_i+z_{i+1}}{2}} |\omega - z_i| \, dF(\omega) \tag{33}$$

with $z_0 = -\infty$ and $z_{k+1} = \infty$. Once the continuous *pdf* for the underlying asset prices is defined, we are able to approximate it obtaining the mass points $z_1, \ldots, z_k$ which minimize the approximation error.

[7] In the case the cost function is Lipschitz constant of order p for some $p > 1$, then the minimization can still be approximated by the Wasserstein distance d_1, but the cost function has to be changed with a non linear transformation in $\omega \mapsto \xi_{1/p}(\omega)$, with

$$\xi_{1/p}(\omega) = \begin{cases} \omega & |u| \leq 1 \\ |\omega|^p \, sgn(p) & |u| > 1. \end{cases}$$

It is therefore equivalent to minimize $d_p(F, \tilde{F})$ or to minimize $d_1(F \circ \xi_{1/p}, \tilde{F} \circ \xi_{1/p})$ with

$$L_p(f) = \inf \left\{ L : |f(u) - f(v)| \leq L |u - v| \max(1, |u|^{p-1}, |v|^{p-1}) \right\}$$

the Lipshitz-constant of order p of f, and

$$d_p(P, \tilde{P}) = \sup \left\{ \int f(\omega) dP(\omega) - \int f(\omega) d\tilde{P}(\omega) : L_p(f) \leq 1 \right\}$$

the Fortet-Mourier distance d_p (see Pflug (2001) for further considerations). For simplicity we assume here that the cost function is Lipschitz constant of order 1.

Using as criteria the minimization of the Wasserstein distance we reduce the simulated density obtained in step 1 in the desired number of scenarios. The simulated density is not a continuous function with a defined functional form. Equation (33) can therefore not directly be applied. A discrete version of (33) can however be formulated.

Suppose we want to reduce the discrete probability density into N points. We can subdivide the density into N sets with $J_i = \left\{\omega : \omega \in \left(\frac{z_{i-1}+z_i}{2}, \frac{z_i+z_{i+1}}{2}\right]\right\}$ for $i = 1, \ldots, N$ and with $z_0 = -\infty$ and $z_{N+1} = \infty$. Equation (33) can then be written in this discrete formulation

$$d_1(P, \tilde{P}) = \sum_{i=1}^{N} \sum_{j \in J_i} |\omega_j - z_i| \, P(\omega_i). \tag{34}$$

We can now reduce the simulated ten thousand point density into N points, by solving equation (34). To solve this discrete formulation of the Wasserstein distance minimization we have, however, to reduce the N mass points z_i with $i = 1, \ldots, N$ to a function of a single variable z. In order to define z we use the following equation as structure for the mass points:

$$z_{\pm l} = m_1(\omega) \pm z \frac{2l}{N} \cdot 3m_2(\omega) \tag{35}$$

where $m_1(\omega)$ and $m_2(\omega)$ are the mean and the standard deviation of the probability density of the asset prices, respectively. N is the total number of mass points we want to obtain, while l represents the points in the left and right region of the density, i.e. $\{\pm 1, \pm 2, \ldots, \pm N/2\}$ if the number of points is even and $\{0, \pm 1, \ldots, \pm(N-1)/2\}$ if the number is odd. For the two extreme points, the fraction $2l/N$ is therefore equal 1 if the number of points is even and $(N-1)/N$ if it is odd. We multiply this fraction by 3 in order to cover the whole density, obtaining a number of mass points which goes from a minimum of $-3z \cdot m_2(\omega)$ to $+3z \cdot m_2(\omega)$. This should avoid excessive concentration of discretizations around the mean and a good covering of the whole density. The obtained N different mass points are therefore a function of mean and standard deviation, and of the optimal value of z which minimizes equation (34). The problem is thus reduced to finding this unique value z which minimizes the Wasserstein distance given the structure of the mass points we have imposed.

The structure of the mass points is symmetric. This seems to be in disagreement with the asymmetry implied in the three GARCH preocesses used to infer the price densities, and probably it is. The problem becomes however irrelevant if we construct multiperiod scenarios. In fact, although the densities are at every investment period discretized using a symmetric structure, over more periods the asymmetric property of the GARCH models prevails. As result we obtain asymmetric scenario trees. Obviously another structure could be defined, and maybe it would be more convenient to consider also the parameter γ (the parameter which accounts for the asymmetric influence) to define z. We leave, however, this analysis of different mass point structures to future research.

Once the optimal value of z is computed and the N mass points defined, their probability is approximated by the percentage of observations lying in each set J_i for $i = \{1, \ldots, N\}$.

So we have obtained the first period nodes which, given the simulated underlying price densities, minimize the discrete formulation of the Wasserstein distance. In order to compute the next investment period (d=2) nodes, we simulate the underlying prices (without recalibrating the GARCH model) for each of the N nodes found in period 1. To maintain the intertemporal relation among subsequent nodes, we explicitly consider the different d=1 nodes as starting points of the asymmetric-GARCH process. This allows us to condition the d=2 nodes to the previous period nodes. This procedure is repeated until the end of the investment time window (after T days). For the Ritchken and Travor (1999) and Heston and Nandi (2000) models we also use the same procedure under the risk-neutral probability measures.

The algorithm can be synthesized in the following six steps, while a graphical representation of the scenario generation procedure is given in Figure 14.

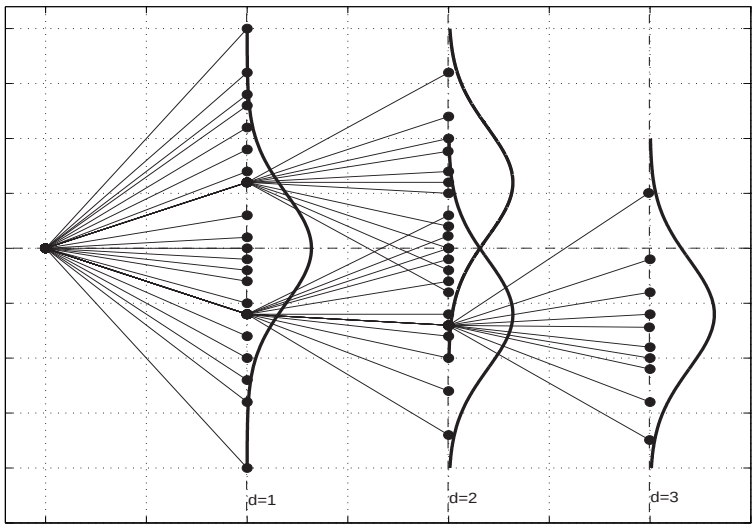

Figure 14: Underlying scenarios generation procedure

1. Estimate the parameters of the process for the underlying asset returns.
2. Simulate the first period (d=1) underlying prices running for (t=ν) days the process in a similar way as in the Barone-Adesi et al.(1999) model, and compute the density function.
3. Solve the Wasserstein minimization and discretize the density function into N points.
4. For each of the found N states, simulate the price density function for the next period, conditioning the density to the previously obtained nodes.
5. Discretize each of the next period probability density functions using the Wasserstein distance minimization criteria.

6. `Repeat point 4 and 5 until reaching day` T `(till the last investment period` d=D`).`

We have thus obtained the entire structure of the underlying scenario tree. For every node we have the price of the underlying asset and the probability of reaching this particular price (for the Ritchken and Travor (1999) and Heston and Nandi (2000) model we also infer the risk-neutral probabilities). Moreover the technique allows us to obtain scenarios which theoretically minimize the approximation error $e(Z, \tilde{Z})$ and are intertemporally consistent in the sense that future periods states directly depend on the states obtained for previous investment periods.

6.2 Option scenario trees

In step 3 of the algorithm the obtained underlying scenario tree is used to infer the prices and probabilities for the options on this asset. First, the same structure (the number of nodes at each investment period) of the underlying scenario tree is maintained for the considered options scenarios. Second, having assumed that all the options mature after T days (D periods), we can use the previously obtained underlying prices for every final leaf to recover the option prices at expiration. These prices are

$$Call_{k_1}(D, n_D) = \max\{S(D, n_D) - X_{k_1}; 0\}$$
$$Put_{k_2}(D, n_D) = \max\{X_{k_2} - S(D, n_D); 0\}$$

for the call and the put, respectively. $S(D, n_D)$ is the underlying price in investment period D if state n_D occurs, with $n_D = 1, ..., N_D^{tot}$ and $N_D^{tot} = N_1 \times N_2 \times \ldots \times N_D$ the total number of nodes in period D. X_k is the option k strike price with $k = 1, ..., nOpt$ and $nOpt$ the total number of options considered[8]. We can now use the option prices at maturity to infer the contingent prices at every node of the tree by discounting the final option values in a risk-neutral world. This approach is well-known when the tree is binomial but is more complex with a multinomial tree. We first have to compute the risk-neutral probabilities of reaching each of the N_D^{tot} final states so that (1) the option theoretical prices obtained discounting backward are as close as possible to the empirical observed market prices in mean-square sense and (2) the possibility of arbitrage is avoided. Second, once the risk-neutral probabilities are computed we have to discount consistently the option prices obtaining the state contingent prices and the relative occurrence probabilities for every node of the tree.

To obtain the risk-neutral probabilities we minimize the mean-square relative difference between the theoretical and the empirical option prices (36) under the constraints that no arbitrage possibility is allowed (37).

$$\min \quad \sum_{k=1}^{nOpt} \left(\frac{\tilde{P}_k - P_k^M}{P_k^M} \right)^2 \tag{36}$$

[8]The total number of options $nOpt$ consists of two groups, with $k_1 = 1, ..., nCall$ and $nCall$ the total number of call options and $k_2 = 1, ..., nPut$, with $nPut$ the total number of put options. $nOpt$ is therefore equal to $nCall + nPut$.

$$\text{s.t.} \quad \begin{pmatrix} 1 \\ S_0 \\ \tilde{P}_1 \\ \vdots \\ \tilde{P}_{nOpt} \end{pmatrix} = \begin{pmatrix} e^{rT} & \cdots & e^{rT} \\ S(D,1)e^{\delta T} & \cdots & S(D,N_D^{tot})e^{\delta T} \\ P_1(D,1) & \cdots & P_1(D,N_D^{tot}) \\ \vdots & \vdots & \vdots \\ P_{nOpt}(D,1) & \cdots & P_{nOpt}(D,N_D^{tot}) \end{pmatrix} \begin{pmatrix} \tilde{\pi}(0,D,1) \\ \tilde{\pi}(0,D,2) \\ \tilde{\pi}(0,D,3) \\ \vdots \\ \tilde{\pi}(0,D,N_D^{tot}) \end{pmatrix} \tag{37}$$
$$\tilde{P}_0^k \geq 0$$
$$\tilde{\pi}(0,D,n_D) \geq 0$$

where $e^{rT}(\tilde{\pi}(0,D,1)),\ldots,e^{rT}(\tilde{\pi}(0,D,N_D^{tot}))$ are the risk-neutral probabilities to reach every final state from date zero (t=0), while r and δ are the risk-free rate and the dividend yield respectively. S_0 is the initial price (at time zero) of the underlying asset, while $\tilde{P}_k$ is the initial option price of the put or the call with strike X_k, respectively.

As result of the minimization we obtain the set of risk-neutral probabilities which ensure that no arbitrage condition is possible, and the today's option prices which are as close as possible to the empirical market option prices in the mean-square sense.

To solve the nonlinear minimization of equation (36) and (37) we need a prior risk-neutral probability set. For the RT and HN GARCH processes the previously obtained risk-neutralized probabilities are used. As explained in section 4, these two models allow us to consistently use the observed risk-free rate r to discount future option prices. For the GJR GARCH process instead, we directly use the objective probabilities. To consistently discount future option prices, we need to recover a discount rate different than those observed empirically. In this case we use the discount rate resulting from the put-call-parity regression of equation (26) on page 47.

Once these risk-neutral probabilities are computed, and given that the structure of the tree and the final leaves option prices are already defined, there is a natural way to recover backward the option prices and the respective risk-neutral probabilities for every intermediate node of the scenario tree. The probability at every node is simply the sum of the risk-neutral probabilities of the subsequent nodes, while the price is

$$P_k(t,n_t) = e^{-r}\left[\sum_{i=1}^{N_{t+1}} \pi(t,t+1,i)\ P_k(t+1,i)\middle| n_t\right], \tag{38}$$

where r is the consistent discount rate, depending on the assumed underlying price process. The option k price at time t and node n_t equals the discounted sum of all subsequent future option prices at period $t+1$ weighted with the respective probabilities, conditional on n_t being the ancestor node. $\pi(t,t+1,n_{t+1})$ is the risk-neutral probability to go from node (t,n_t) to the subsequent node $(t+1,n_{t+1})$, and $P_k(t+1,n_{t+1})$ is the price of an option with strike k at the node $(t+1,n_{t+1})$. Solving the equation for every node and every investment period, we obtain the entire multinomial scenario tree for all the considered options.

7 Empirical analysis

In this section we present the empirical results of the generation of scenario trees for a specific underlying asset and the respective options on this asset. The empirical analysis starts with the description of the data. It proceeds with the estimation of the three GARCH models and the generation of the scenario trees.

7.1 Data description

To compute the underlying asset and the options scenario trees, we consider the German DAX 100 index[9] time series from January 7, 1991 to April 23, 2004. The computed one-month scenarios start on April 23, 2004 when the index price was euro 4103.61. We consider 26 European call and 26 put options on the DAX 100 all maturing one month later (May 21, 2004). The prices (in euro) on April 23, of the 52 options and their strikes are reported in Table 10.

Strike	Call Price	Put Price	Strike	Call Price	Put Price
3300	814.4	1.2	3950	196.3	32.0
3350	764.9	1.6	4000	157.0	42.7
3400	715.4	2.0	4050	121.2	56.8
3450	665.8	2.4	4100	88.8	74.4
3500	616.3	2.8	4150	62.9	98.3
3550	567.2	3.7	4200	41.5	126.9
3600	518.6	4.9	4250	25.7	161.0
3650	469.8	6.1	4300	16.2	201.4
3700	421.6	7.8	4350	9.5	244.7
3750	374.6	10.6	4400	5.3	290.4
3800	327.6	13.6	4450	3.3	338.3
3850	282.4	18.3	4500	1.9	386.8
3900	238.5	24.4	4550	1.3	436.1

Table 10: Prices and strikes (on April 23, 2004) for the 26 call and 26 put options on the DAX 100 maturing after one month (May 21, 2004).

We consider all the options present in the market, without excluding any deep out of the money options. As many of the stocks in the DAX 100 pay dividends, one needs a time series of dividends for the index. We compute it by summing up the single DAX 100 stock daily dividends, weighted by their capitalization. We then subtract the dividend time series from the current index level. This should allow us to not consider the dividend yield δ in the option valuation formula of equation (36). For the RT and HN models, we use the historical risk-free rate over the option time to expiration period, i.e. the average of the one month euribor from December 30, 1998 to April 23, 2004. The resulting annual risk-free rate is about 3.32%. For the GJR model we compute the risk free rate using the

[9]The choice of this asset is purely for convenience, since the data of the index and the options were already set. With regard to this I thank my brother Andrea Laurent.

put-call parity formula (26). We regress the difference of the 26 call and put option prices on April 23, 2004 over the index level and the 26 different strike prices.

$$c_t - p_t = S_t e^{-q(T-t)} - X e^{-r(T-t)}.$$

If we do not impose that the dividend yield be equal to zero (using directly the index level, without any depuration from the dividends), we obtain an annual risk-free rate of about 1.10% and an annualized dividend yield of 1.015%. Using a restricted least squares method (and the index price depurated from the dividends) and imposing the dividend yield to be equal zero, we obtain an annual risk free rate of about 1.56%. Graphically the restricted least-square regression is represented in Figure 15. The time series come from the Bloomberg (all the prices) and DataStream (dividends and capitalizations) dataset.

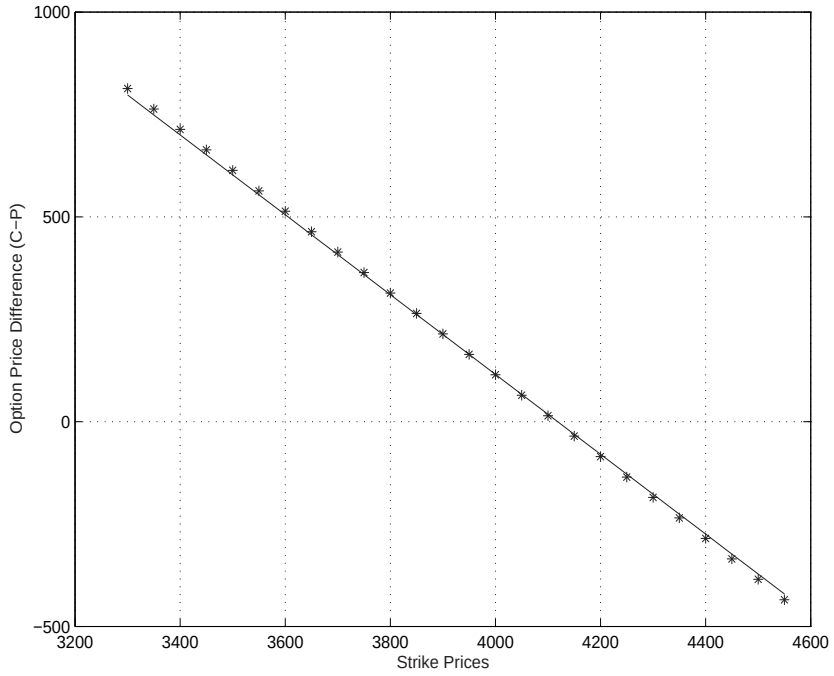

Figure 15: DAX 100 put-call parity restricted least square regression used for the determination of the consistent discount rate

7.2 Parameter estimation

We estimate the three GARCH processes, using the period going from August 23, 1996 to April 23, 2004 (2000 observations) for the estimation of the process parameters. This number of around eight years of data seems to be a good trade-off to obtain parameters which satisfy all the stationarity conditions as well as describe well the last years of assets returns used to simulate future underlying price densities. Unlike continuous-time stochastic volatility models in which the volatility process is unobservable, all the parameters in our valuation formula can be easily estimated from history of asset prices. We do this with maximum likelihood estimation (MLE). The values of the parameters, their standard error and t-statistic under the objective probability measure, as well as

the value of the log-likelihood function are reported in Table 11. The current April 23, 2004 estimates on a daily base of the volatility h_t and the random variable v_t are 0.010 and 0.98 for the RT; 0.012 and 0.92 for the HN; and 0.011 and 1.03 for the GJR process respectively and will be used as starting values to simulate the future underlying asset price density functions. The skewness parameter γ is positive and significant for all the three processes. This indicates the presence of the empirically observed "leverage effect", i.e. the strong negative correlation between shocks to returns and volatility. Also the GARCH parameter β is highly significant for all the three processes, and is a symptom of high volatility clustering. The t-statistic goes from 37.27 for the RT model to 98.32 for the GJR model. Unlike the α coefficients and the constant in the volatility function of the RT GARCH process, the other parameters do not seem to be significantly different from zero. In particular the volatility risk premium λ does not seem significant either for the RT or for the HN process. Computing the annualized (252 days) long run volatility (θ) implied by the parameters estimates, we obtain similar results for all the processes.[10] For the RT model we obtain an annualized long-term volatility equal to $\theta_{RT} = 27.38\%$, for the HN, $\theta_{HN} = 26.67\%$, while for the GJR, $\theta_{GJR} = 27.39\%$.

Finally, the values of the log-likelihood function (LLF) also, do not differ so much across the three models. These are 5524.9 for the RT, 5501.6 for the HN, and 5535.1 for the GJR model, respectively. Moreover, The Akaike (AIC) and Schwarz (BIC) information criteria are equal to -5.52 and -5.51 for the RT, -5.50 and -5.49 for the HN and -5.53 and -5.52 for the GJR model, respectively.

Figures 16 to 18 show the series of historical returns, standardized innovation, and standard deviation for the three models. While the standardized innovations are almost identical for the three models, the HN GARCH shows lower standard deviation than the other two processes. Although quite similar, the RT GARCH shows higher peak in standard deviation than the GJR model. Since the filtered historical simulation can be applied only if the standardized residual returns $e_t = \hat{\varepsilon}/\sqrt{\hat{h}_t}$ are i.i.d. we compute for all the three standardized innovation time series the Ljung-Box test. For none of the considered lags, from 1 to 36, we can reject the null hypothesis that the autocorrelation is statistically equal to zero. This suggests that the obtained standardized residuals are effectively independent and probable suitable for simulation. We can therefore proceed with the simulation of the underlying price probability densities using the Barone-Adesi, Giannopoulos, and Vosper (1999) method.

[10] For the HN model the annualized long run volatility is

$$\theta_{HN} = \sqrt{\frac{252(\omega + \alpha)}{(1 - \beta - \alpha\gamma^2)}}$$

(see Heston and Nandi (2000), footnote on Table 1, page 597). For the RT model we define the long run annualized volatility to be

$$\theta_{RT} = \sqrt{\frac{252\omega}{(1 - \beta - \alpha - \alpha\gamma^2)}}$$

while for the GJR model

$$\theta_{GJR} = \sqrt{\frac{252\omega}{(1 - \beta - \alpha - \gamma/2)}}.$$

$$r_t = r_f + \lambda\sqrt{h_t} - \tfrac{1}{2}h_t + \sqrt{h_t}v_t$$
$$h_t = \omega + \beta h_{t-1} + \alpha h_{t-1}(v_{t-1} - \gamma)^2$$

	Coefficient	Std. Error	t-stat.
λ	0.023985	0.032645	0.734724
ω	5.25E-06	1.23E-06	4.274531
α	0.068634	0.008911	7.707490
β	0.870989	0.019250	45.24530
γ	0.789066	0.153477	5.141275
LLF			5524.9

$$r_t = r_f + \lambda h_t + \sqrt{h_t}v_t$$
$$h_t = \omega + \beta h_{t-1} + \alpha(v_{t-1} - \gamma\sqrt{h_{t-1}})^2$$

	Coefficient	Std. Error	t-stat.
λ	0.520339	0.525809	0.989597
ω	7.77E-11	4.14E-07	1.87E-04
α	1.10E-05	1.74E-06	6.295519
β	0.883161	0.023697	37.26885
γ	84.11185	17.17399	4.897632
LLF			5501.6

$$r_t = \mu + \varepsilon_t$$
$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma I_{t-1}\varepsilon_{t-1}^2.$$

	Coefficient	Std. Error	t-stat.
ω	3.90E-06	8.09E-07	2.389324
α	0.017221	0.009895	3.939310
β	0.918666	0.009344	98.31835
γ	0.102024	0.014308	7.130817
LLF			5535.1

Table 11: Coefficient of the 3 GARCH processes (RT, HN, GJR) for the DAX 100 (01.07.1991 - 04.23.2004 period, 3370 observations) under the objective probability measure.

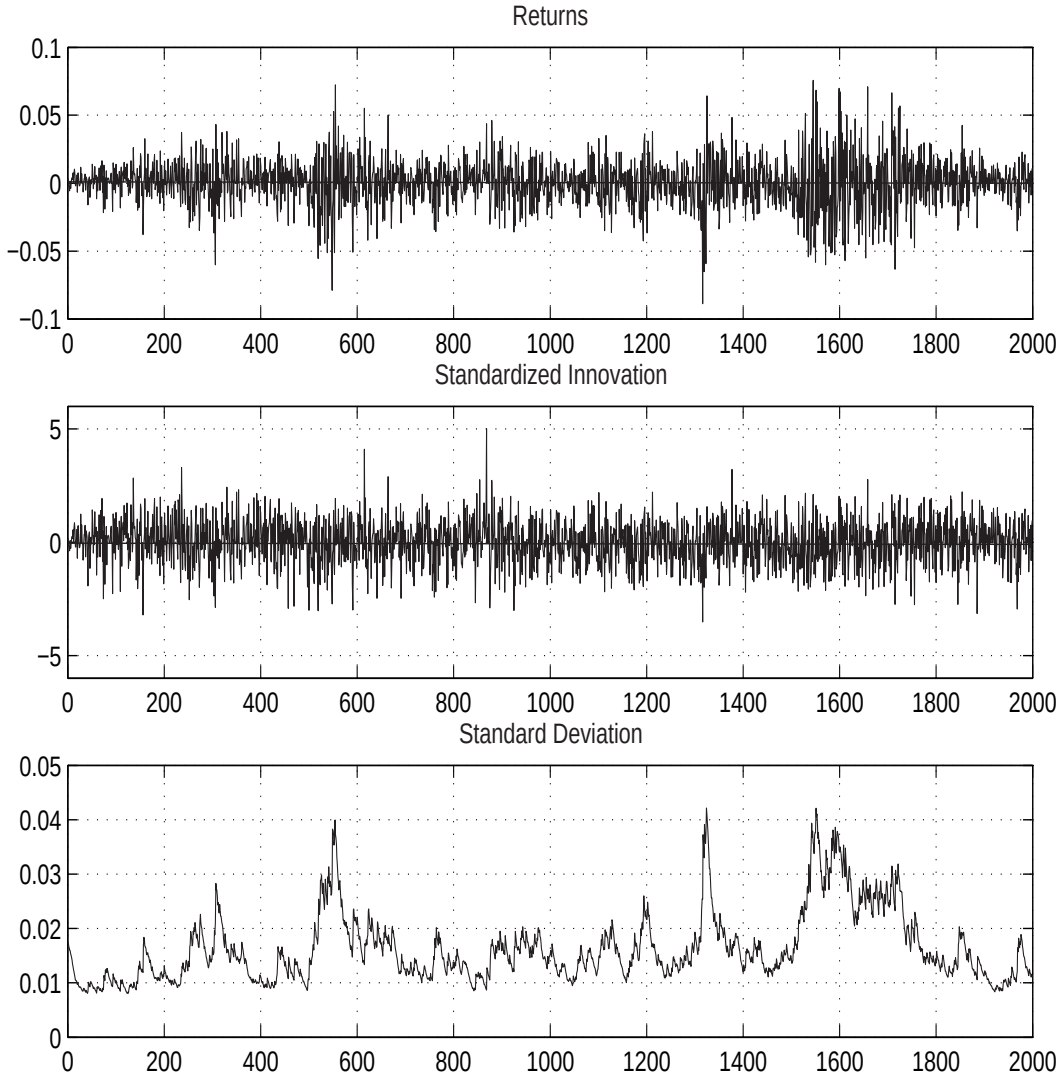

Figure 16: estimated series of returns (r_t), innovation (ϵ_t), and standard deviation (h_t) for the DAX 100 over the 08.23.1996 - 04.23.2004 period for the Ritchken and Trevor (1999) model.

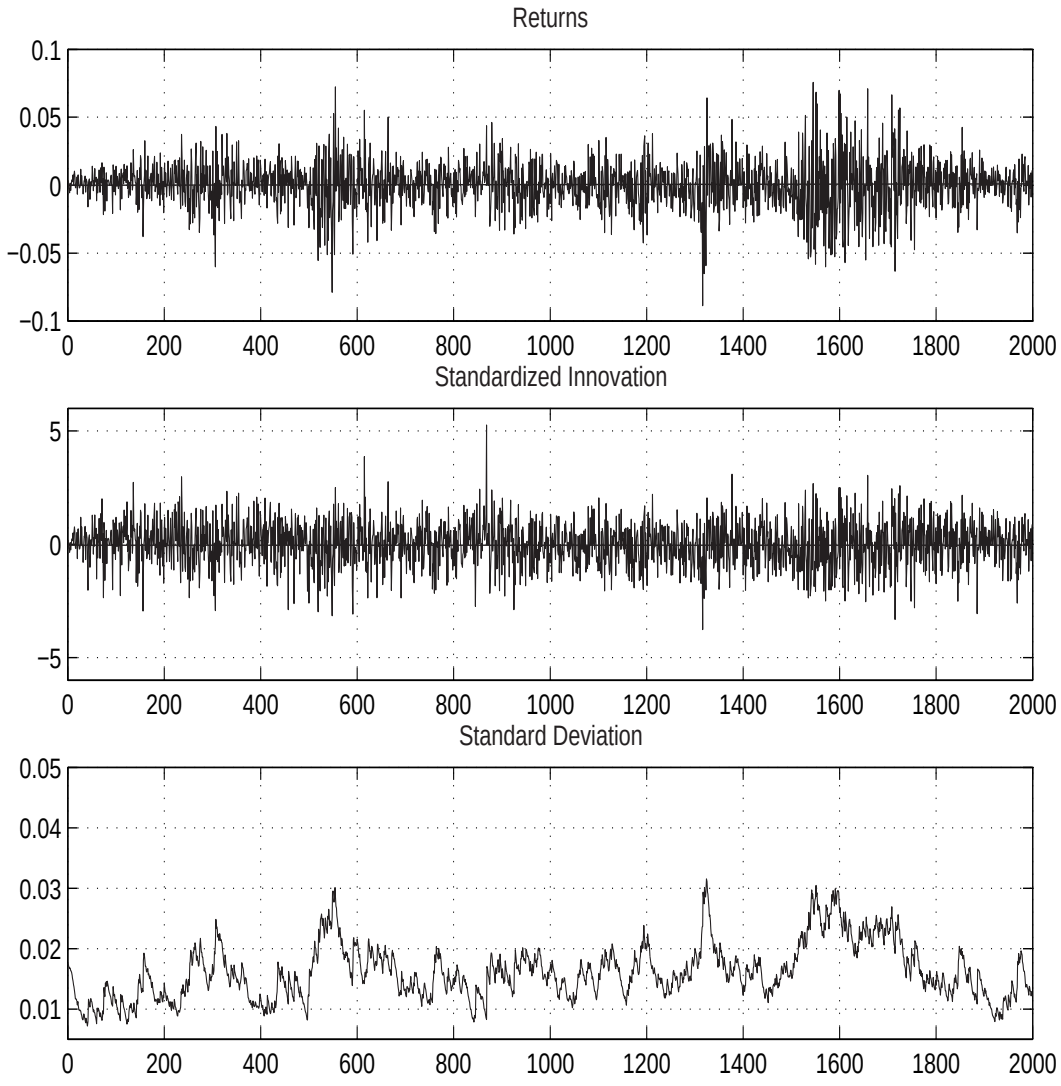

Figure 17: estimated series of returns (r_t), innovation (ϵ_t), and standard deviation (h_t) for the DAX 100 over the 08.23.1996 - 04.23.2004 period for the Heston and Nandi (2000) model.

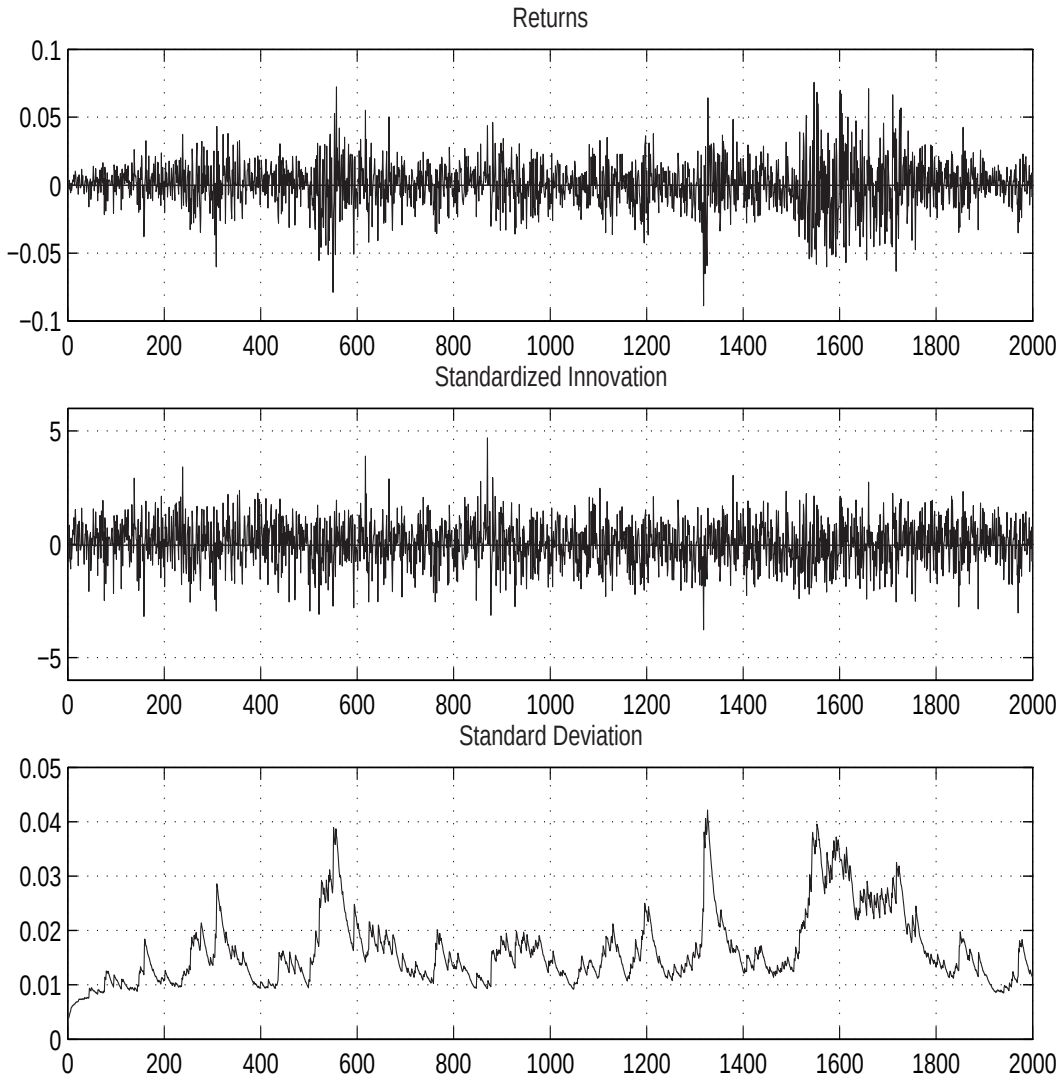

Figure 18: estimated series of returns (r_t), innovation (ϵ_t), and standard deviation (h_t) for the DAX 100 over the 08.23.1996 - 04.23.2004 period for the Glosten, Jagannathan and Runkle (1993) threshold GARCH model.

7.3 Scenario Generation

Once the process parameters are estimated, we start with the computation of the four weeks scenario tree for the underlying asset under the three different GARCH processes. We infer the first week following April 23, 2004 probability density using the methodology based on the Barone-Adesi, Giannopoulos, and Vosper (1999) method, explained in section 4. We draw from the last three years of data ten thousand standardized residuals with replacement and compute the next five days future prices from the respective GARCH process. We then discretize the ten thousand points density using the discrete form of the Wasserstein distance minimization of equation (34) on page 61. For each of the obtained points we repeat the simulation and the discretization again obtaining for week two a limited number of nodes intertemporally connected with the previous one. We continue the procedure until week four (May 21, 2004), the option expiration date. The result is a four period scenario tree, for which we have chosen a number of nodes equal 15 for the first, 9 for the second and 5 for the third and fourth week. This gives us a total number of $15 \times 9 \times 5 \times 5 = 3375$ different scenarios for the underlying prices. The computational time to recover the scenario trees using a Pentium 4 with 3 GHz, is about 40 minutes for the Ritchken and Travor (1999) and the Heston and Nandi (2000) GARCH processes (for which we compute contemporaneously the scenarios under both, the objective and the risk-neutral probability measures) and about 20 minutes for the Glosten, Jagannathan and Runkle (1993) threshold GARCH process (for which we compute only the scenarios under the objective probability measure). This time seems quite reasonable considering the large amount of final scenarios, and could probably be reduced by parallelizing the procedure.

The three scenario trees under the objective probability measure are graphically reported in Figures (19) to (21). As explained in section 6.1, although the discretization procedure is based on a symmetric structure, the scenario trees are not symmetric. For all the three processes, in period four the number of scenarios showing a negative return (i.e. a price below the starting one) exceeds the number of positive scenarios. Especially the GJR tree looks highly skewed with a large number of extremely negative nodes. This is due to the structure of the process and the high significance of the skewness parameter γ. Second, instead that the symmetric structure of the approximation procedure brings the first period nodes to be perfectly symmetric. For a better description of the first period nodes it would be probably more appropriate to consider the parameter γ in the structure of the mass points of equation (35) on page 61. Finally, the scenario tree related to the HN process is for each of the four considered weeks following April 23 more spread out than the one related with the RT process.

To get a general idea about the quality of our scenarios, we compare the simulated scenarios of the returns with the historical scenarios obtained computing the DAX 100 returns for all overlapping four weeks over the last three years (a total number of 746 scenarios). In Figure (22) are reported the empirically observed scenarios and the simulated upper and lower limit scenarios for each of the three different processes. Compared with the empirical returns, the simulated nodes seem too concentrated around zero in the first investment period while they seem too spread out for the last period (at least for the negative side). In fact, unless for the first (and one case for the second) week as well as for the positive side of the GJR process, the upper and lower limit scenarios seem too far

from the limits of the observed empirical scenarios, especially for the negative side. The concentration in the first period scenarios is due to the approximation procedure. For the first investment period we simulate only one probability density. Discretizing this density we inevitably obtain a set of mass points which is narrower than the original density. The spread of the last period could instead be explained by the difference in the number of scenarios. While with the simulation we compute 3375 final possible returns, the empirical scenarios are only 746. The simulated scenarios could therefore describe events which do not enter in the last three years of observed data. Moreover, the graphically reported simulated scenarios represent only the possible returns and do not consider the probabilities of reaching each node. We should weight the different node returns with the related occurrence probabilities. Figure (23) reports the three simulated and the historical density functions of the four weeks asset return scenarios. The densities of the scenarios related to the RT and HN processes are effectively characterized by a higher variance compared to the historical distribution, while for the threshold GARCH of Glosten, Jagannathan, and Runkle (1993) the scenarios are more concentrated around the mean, and show a lower dispersion than the other two processes and the historical density function (especially for the positive side).

The Ritchken and Travor (1999) GARCH process seems the one which better describes historical returns. Especially around the mean, the RT process seems to be superior to the HN and GJR processes, which under- and overestimate the probability of the scenarios, respectively. The relative closeness between the RN and the historical probability density function makes us quite confident about the quality of the generated scenarios.

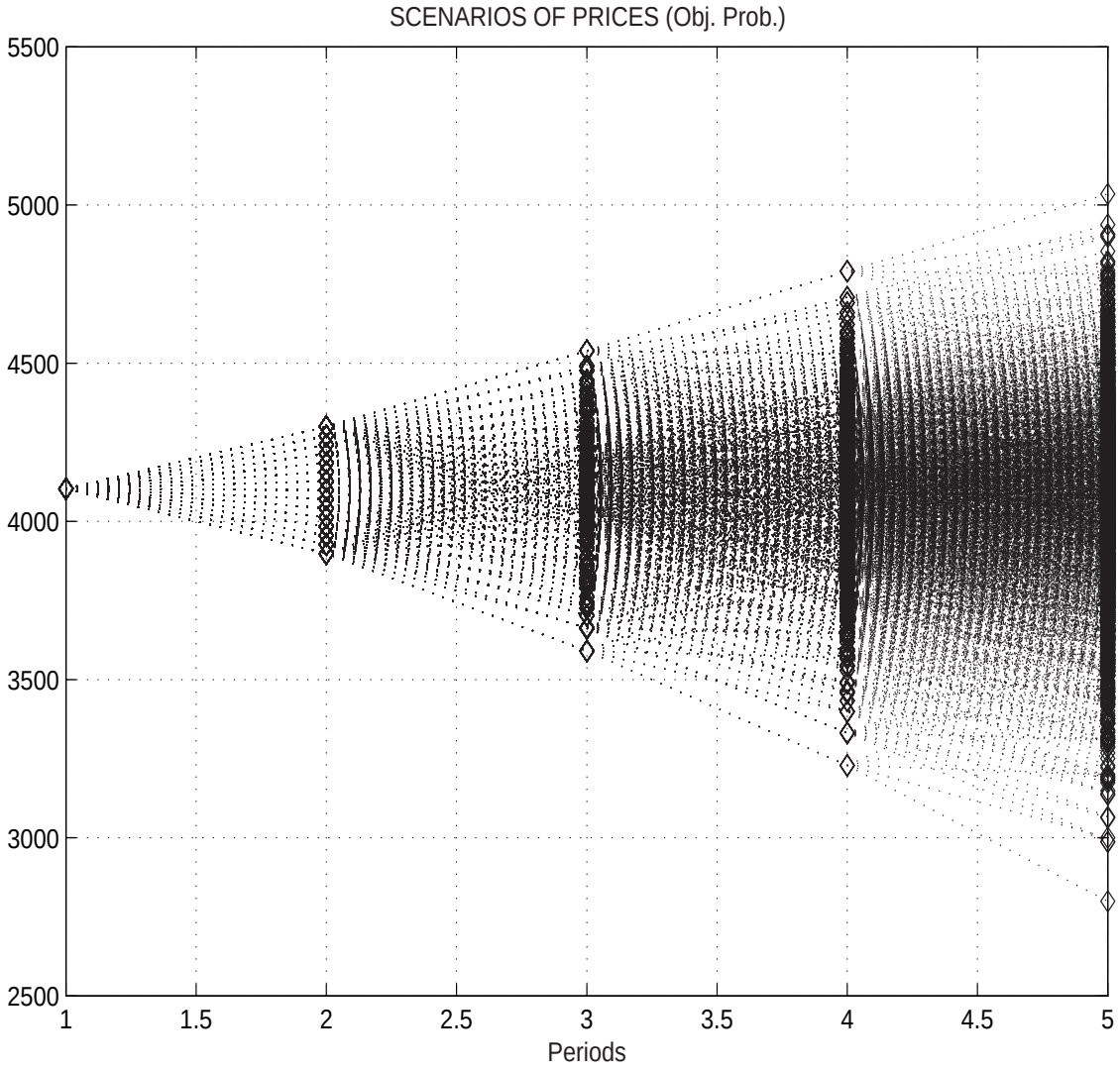

Figure 19: $15 \times 9 \times 5 \times 5$ scenario tree for the DAX 100 prices (Ritchken and Travor (1999) GARCH process).

The figure reports the simulated four weeks scenario tree of the DAX 100 prices starting at April 23, 2004 (when the price was equal to 4103,61). The process used to simulate the future prices is the Ritchken and Travor (1999) GARCH.

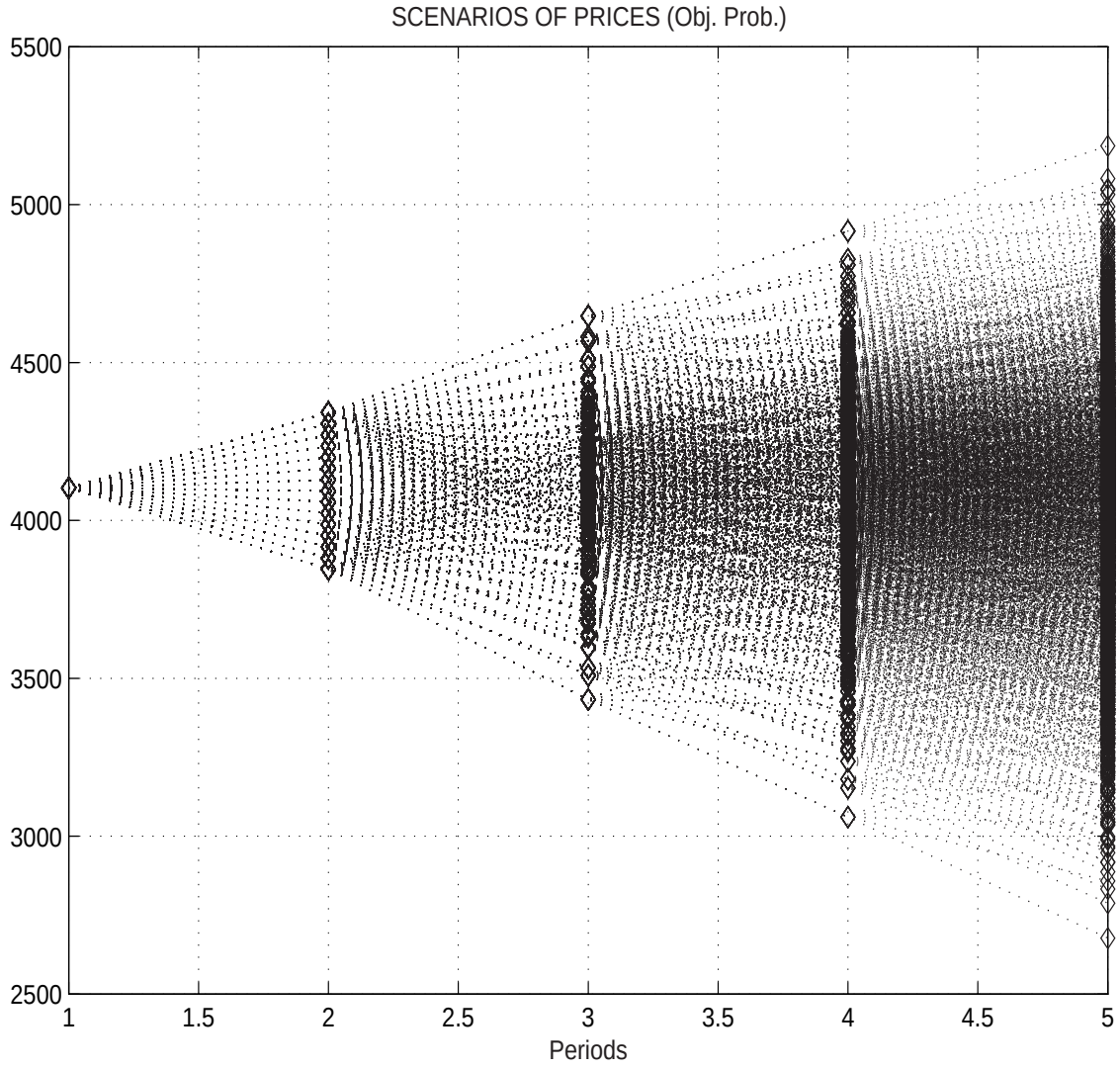

Figure 20: $15 \times 9 \times 5 \times 5$ scenario tree for the DAX 100 prices (Heston and Nandi (2000) GARCH process).

The figure reports the simulated four weeks scenario tree of the DAX 100 prices starting at April 23, 2004 (when the price was equal to 4103,61). The process used to simulate the future prices is the Heston and Nandi (2000) GARCH.

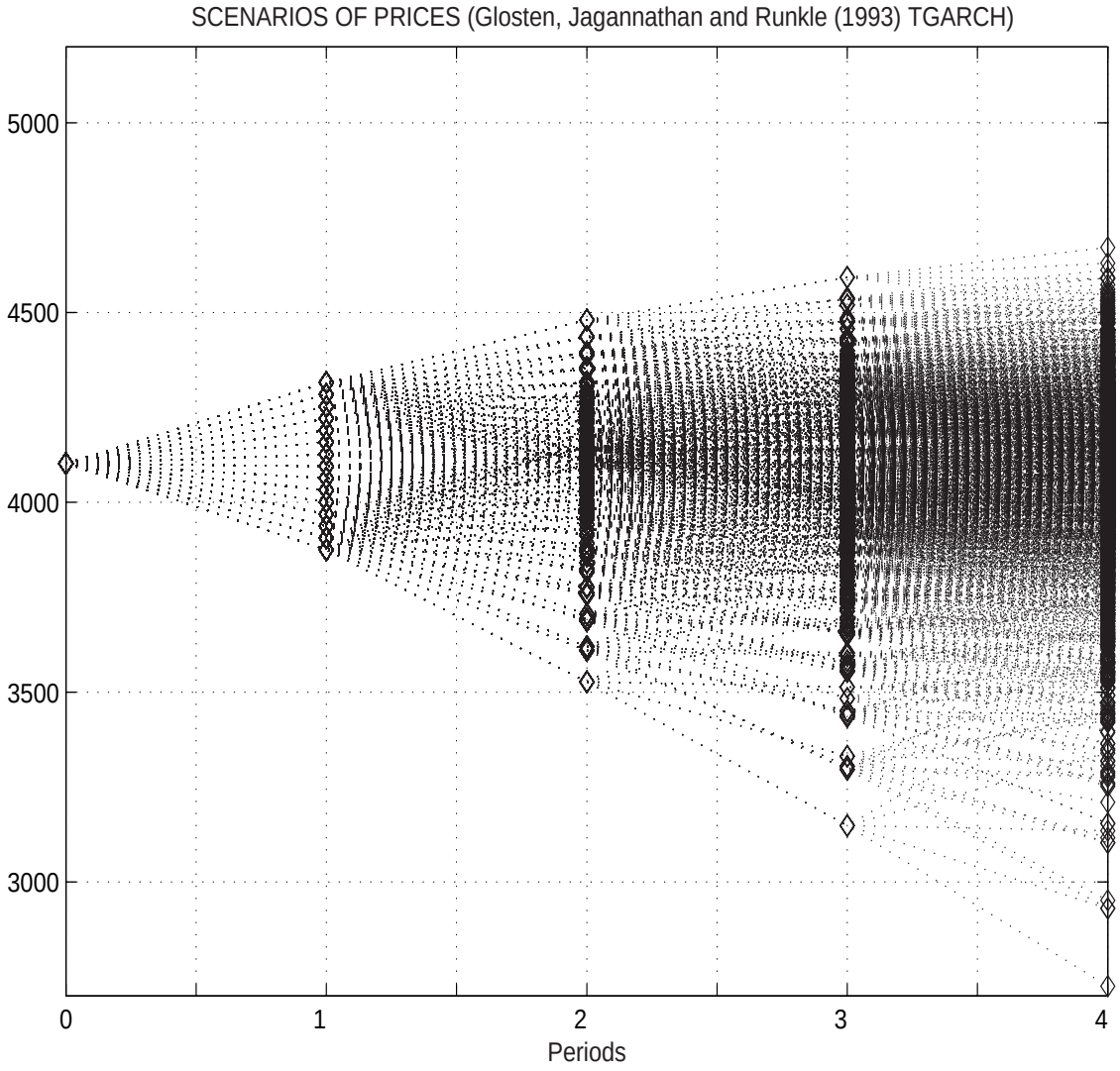

Figure 21: $15 \times 9 \times 5 \times 5$ scenario tree for the DAX 100 prices (Glosten, Jagannathan and Runkle (1993) TGARCH process).

The figure reports the simulated four weeks scenario tree of the DAX 100 prices starting at April 23, 2004 (when the price was equal to 4103,61). The process used to simulate the future prices is the Glosten, Jagannathan and Runkle (1993) TGARCH.

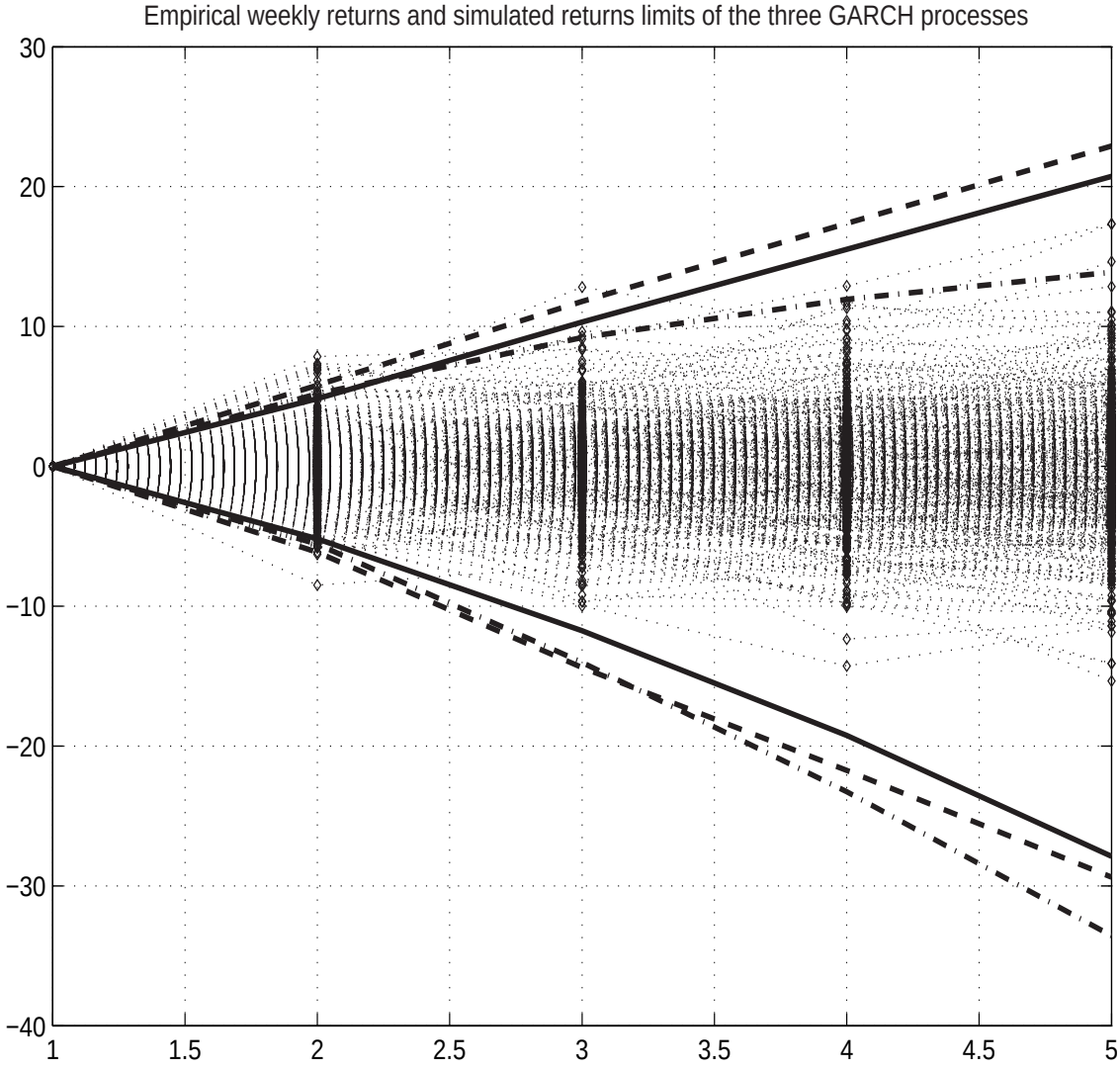

Figure 22: Empirical 4 weeks returns scenarios and upper and lower limits of the three simulated 4 weeks scenarios (RT: — ; HN: - - ; GJR: - · -)

Simulated upper and lower limits of the DAX 100 returns starting on April 23, 2004 for all the three considered processes. The dotted lines with ⋄ are the observed four week empirical scenarios over the period from June 11, 2001 to April 23, 2004 (746 scenarios).

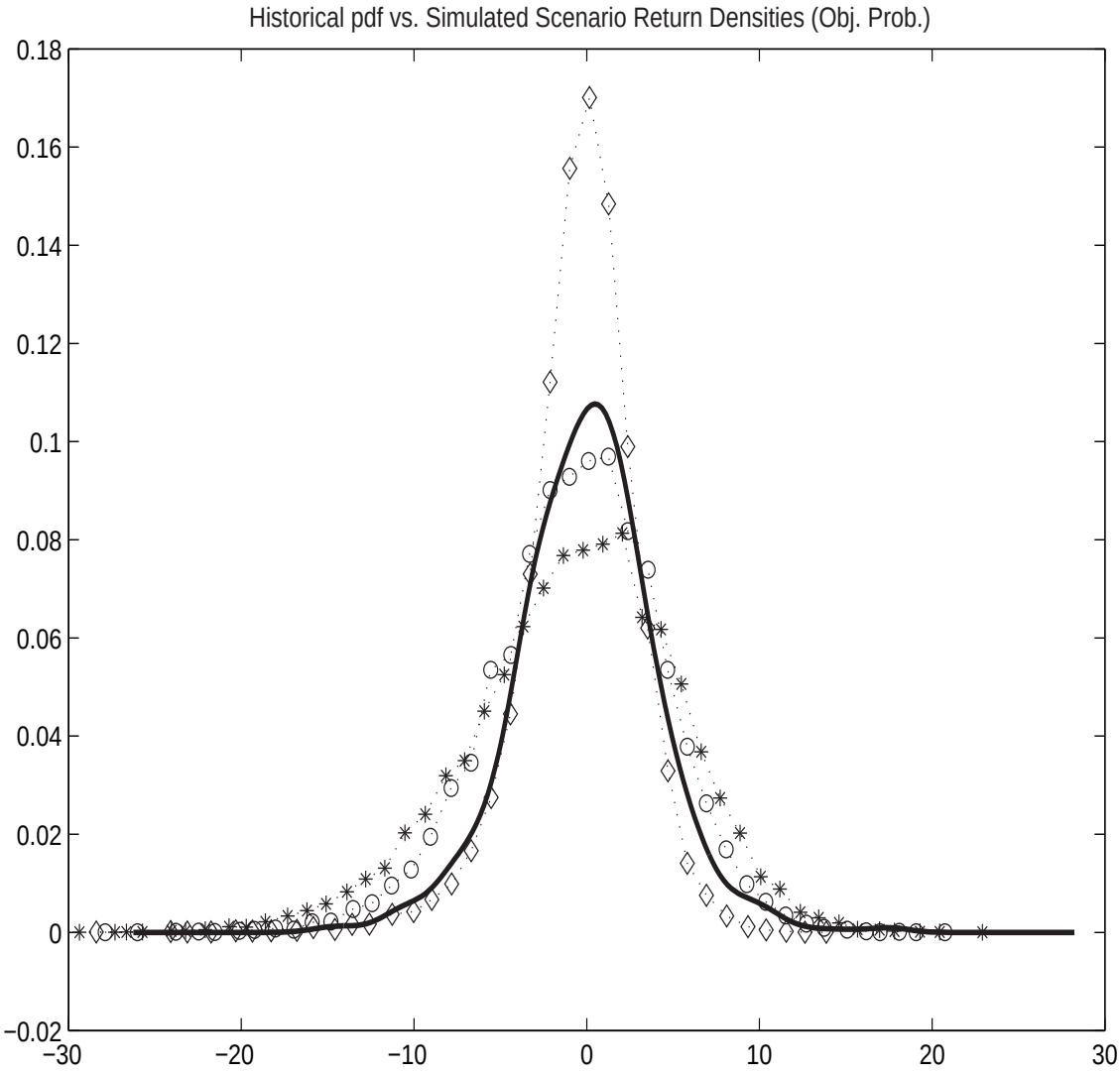

Figure 23: Empirical (continued line) vs. simulated (dotted line, ∘: RT-GARCH process, ∗: HN-GARCH process, and ⋄: GJR threshold GARCH process) density.

Empirical observed probability density function of the four week returns over the period from June 11, 2001 to April 23, 2004 (750 observations) and the simulated week four scenarios (May 21, 2004) weighted with the occurrence probabilities.

7.4 Option price scenarios

Once we have obtained the underlying scenario tree, we can infer the scenarios for all the considered 52 option prices (26 calls and 26 puts). Since the end of the scenario tree corresponds to the time of option expiration, we can use the final leaves DAX 100 prices to directly compute the option prices at expiration equal to $C_k(H, n_H) = \max\{S(H, n_H) - X_k; 0\}$ for the call options and $P_k(H, n_H) = \max\{X_k - S(H, n_H); 0\}$ for the put options, respectively. For the RT and HN processes we have also already computed the underlying price scenarios under the risk neutralized probability measure using equation (23) and (24). Theoretically knowing the option payoff at expiration and the risk neutral probabilities, we could directly compute the option prices for each of the node of the scenario tree, discounting the final payoff in a similar way as in Cox, Ross, and Rubinstein (1979) for a binomial lattice. Because of the historical nature of the information embedded in the obtained underlying prices, however, this could be far from an optimal solution. The forward-looking nature of the information in option prices should be considered in the computation of future contingent prices. For this reason we use the obtained underlying risk-neutral probabilities only as priors in the minimization of equation (36) on page 63 under the no-arbitrage conditions (37) and the two positively constraints $\tilde{C}_0^k \geq 0$ and $\tilde{\pi}(0, H, n_H) \geq 0$. This allows us to combine both the information from the historical underlying prices as well as the forward-looking information of the observable option prices. For the risk-free rate we use the average historical rate observed in the market (3.32%). What we obtain are the ending risk-neutral probabilities, i.e. the probabilities to go from the 52 option prices on April 23 to the 52×3375 prices at expiration (May 21), which in the mean square sense minimize the difference between the simulated option prices and the prices observed in the market.

For the GJR threshold GARCH we use the underlying prices under the objective probability as prior for the minimization. As the risk-free rate we use the result from the restricted least square regression of the difference between the put and call prices, over the strike prices and the underlying price, as explained in section 6. The resulting annual risk-free rate is about 1.56%.

With the obtained ending probabilities, we can infer the option prices and the probabilities for any node of the scenario tree, working backward, i.e. for any strike price we can compute the price of the related put and call options maturing in period D at every node of the tree using equation (38). The option price percentage errors for each of the three processes are

$$100\left(\frac{C_0^k - C_M^k}{C_M^k}\right)$$

and are reported in Figure (24), while Figure (26) shows the simulated and the empirical prices as function of the strike prices. C_0^k is the April 23 simulated price, while C_M^k is the observed option market price, of an option maturing at period D (May 21, 2004). For the RT and HN processes we obtain, although starting from different priors, exactly the same option prices. This makes us confident about the robustness of the methodology. The maximal percentage error for these two processes is in absolute values about 2.7% for both, the call and the put option prices, respectively. This occurs for the two at-the-value options. For the GJR process, the absolute maximal error is much higher and about

46.9% for the call and 3.71% for the put.[11] For 18 out of 52 options using the RT and HN processes, the percentage error is less than 1% while for 40 option prices this error is less than 2%. For the GJR threshold GARCH 8 and 31 percentage errors are less than 1% and 2%, respectively. As an overall measure of the quality of the calibration we compute the average pricing error (ape) with respect to the mean price,

$$ape := \frac{\sum_{k=1}^{nOpt} \left| \tilde{C}_k - C_k^M \right|}{\sum_{k=1}^{nOpt} C_k^M}.$$

The average pricing error for the Ritchken and Travor and for the Heston and Nandi processes is 1.11%, while for the Glosten, Jagannathan and Runkle process it is 2.02%. The overall pricing performance is quite satisfactory given the wide range of strikes of the options used for the calibration.

Moreover, to analyze if a process specification is superior to the others we test if the absolute error between the theoretical and the empirical observed prices is statistically different for the three different specifications. Since the prices obtained with the RT and HN processes are exactly the same, we actually test the results obtained with these two GARCH models against those obtained with the GJR process. The test is defined as follows. Let $\tilde{U}_k$ be the absolute error at every strike level k, i.e.

$$\tilde{U}_k = \left| \tilde{C}_k - C_k^M \right|.$$

The realized error difference at strike k between the RT (HN) precess and the GJR specification is

$$\hat{D}_k = \tilde{U}_{k;RT} - \tilde{U}_{k;GJR} \ .$$

We test the null hypothesis that the differences $\hat{D}_k$ have mean zero against the alternative of mean less than zero, i.e. the hypothesis that the average errors implied by the RT (HN) specification are smaller than those using the GJR GARCH process. The t-type test statistic is

$$\sqrt{nOpt}\frac{\bar{D}}{\hat{\sigma}_{D;\infty}}, \text{ where } \bar{D} = \frac{1}{nOpt}\sum_{k=1}^{nOpt} \hat{D}_k,$$

where $\hat{\sigma}_{D,\infty} = (2\pi)\hat{f}_{\hat{D}}(0)$ and $\hat{f}_{\hat{D}}(0)$ a smoother periodogram estimate at frequency zero based on $\hat{D}_1, \ldots, \hat{D}_{nOpt}$. Then, under the given null hypothesis

$$\sqrt{nOpt}\frac{\bar{D}}{\hat{\sigma}_{D;\infty}} \Longrightarrow N(0,1) \ \ (nOpt \to \infty)$$

where

$$\sigma^2_{D;\infty} = \sum_{j=-\infty}^{+\infty} Cov[\hat{D}_0, \hat{D}_j] = 2\pi f_{\hat{D}}(0),$$

[11] In Figure (24) the two extreme call option errors (11.6% and 46.9%) are outside the margins of the graph.

and $f_{\hat{D}}(0)$ is the spectral density of $\{\hat{D}_k\}_k$ at zero.

We choose three different samples for the test. The first considers the whole set of 52 options, while the second and the third consider only the 26 Call and 26 Put, respectively. For the three samples none of the differences between the pricing errors using the a RT (HN) or a GJR specifications, are statistically different from zero. We conclude that it is not possible to assert that one GARCH specification is statistically better than another to price options with our algorithm.

Finally Figure (27) reports the three objective and risk-neutral probability density functions. Note that although the Ritchken and Travor and the Heston and Nandi processes impose that the volatility of the risk-neutral and the objective distributions be the same, the calibration procedure on option prices makes us relax this assumption. In fact, the obtained objective and risk-neutral probability density functions differ not only in the drift but also in the volatility.

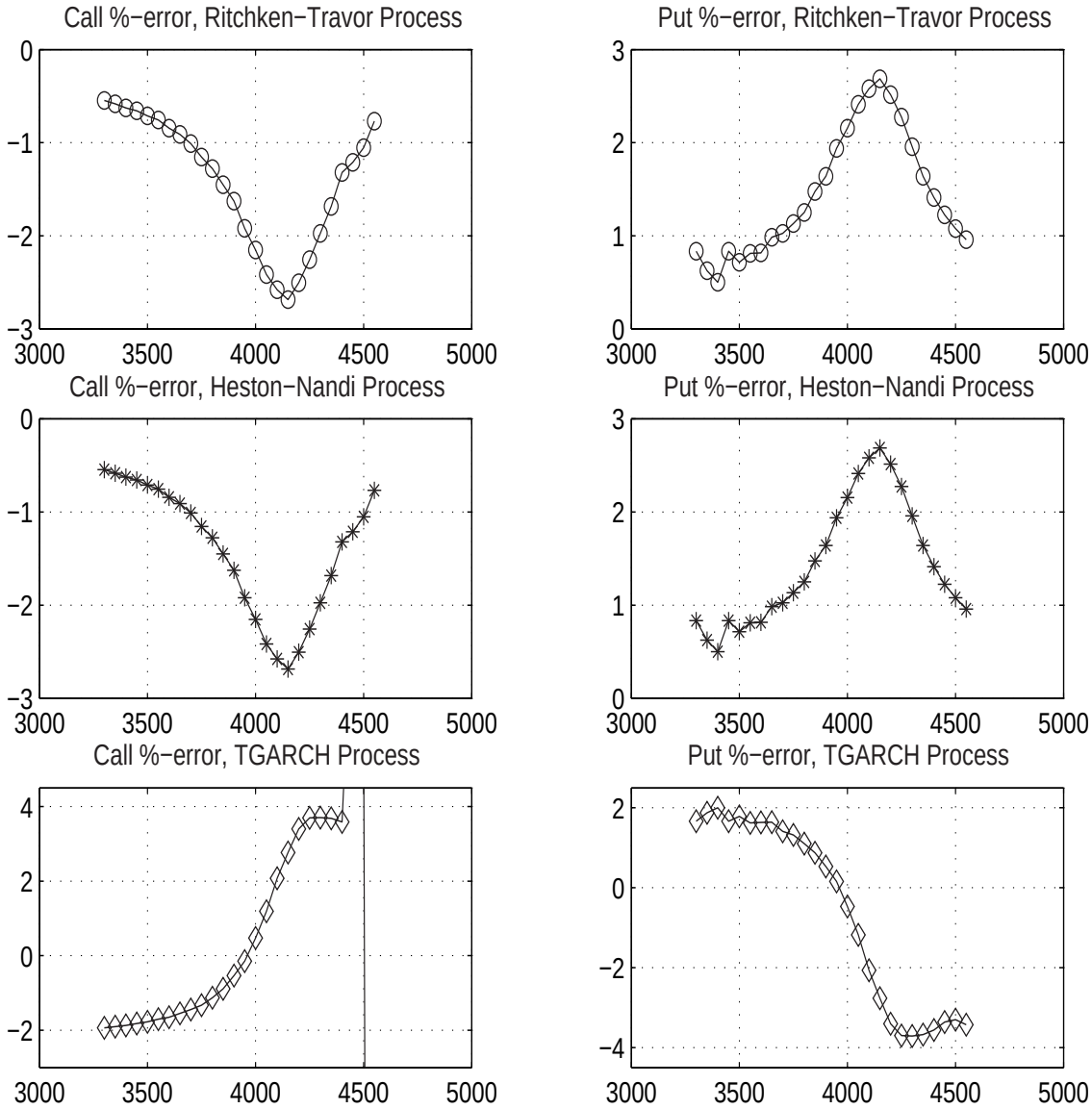

Figure 24: Percentage error between simulated and empirical call and put option prices on April 23, 2004

The six graphs report the percentage error $100\left(\frac{C_0^k - C_M^k}{C_M^k}\right)$ as function of the strike prices. For the GJR TGARCH the two extreme call option price errors (11.6 % and 46.9%) are outside the margin of the graph.

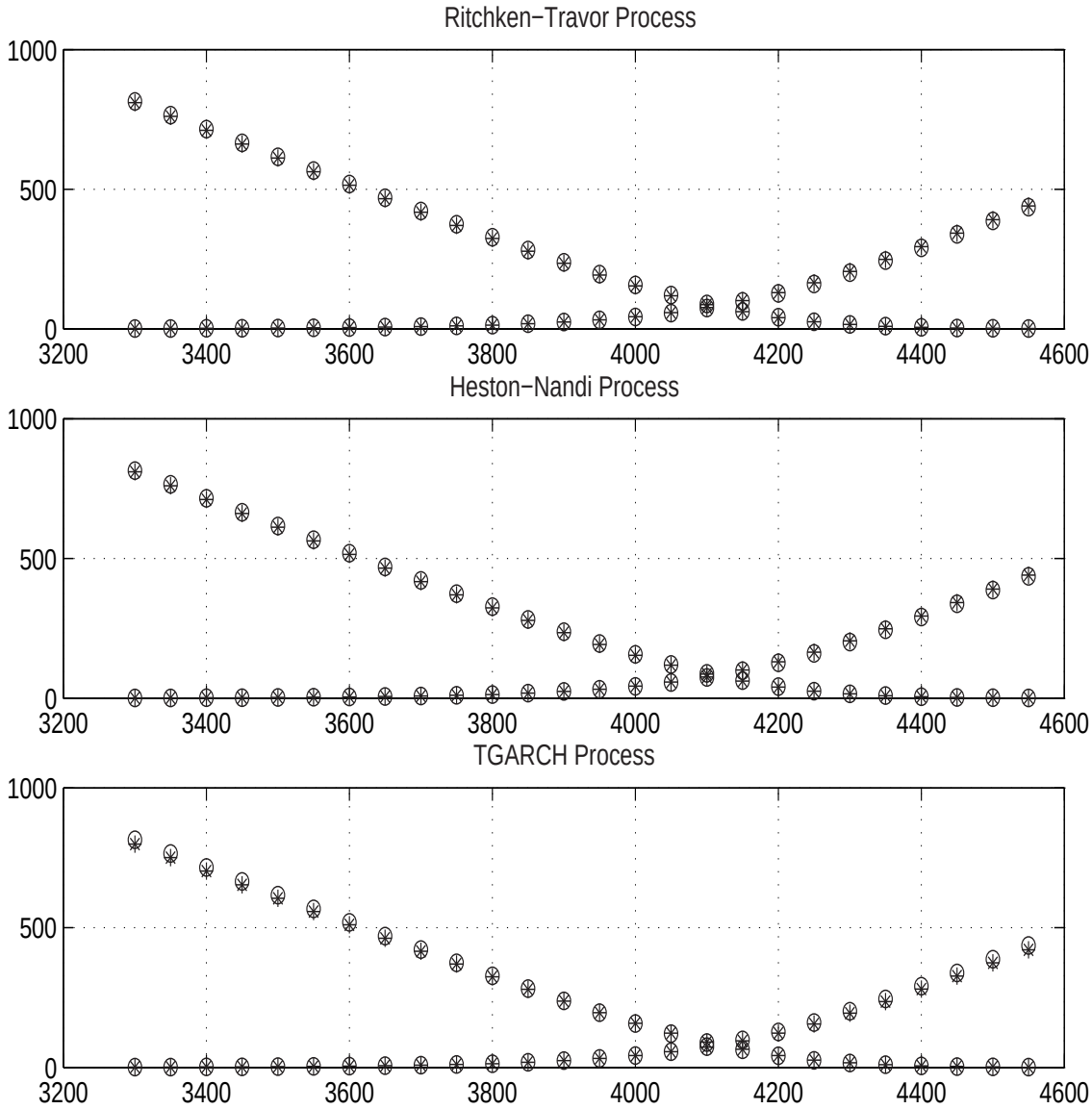

Figure 25: Empirical ($\circ$) and simulated ($*$) call and put option prices

Empirical observed DAX 100 call and put option prices on April 23, 2004 (Maturity equal to May 20, 2004) and simulated option prices computed using the three different GARCH processes

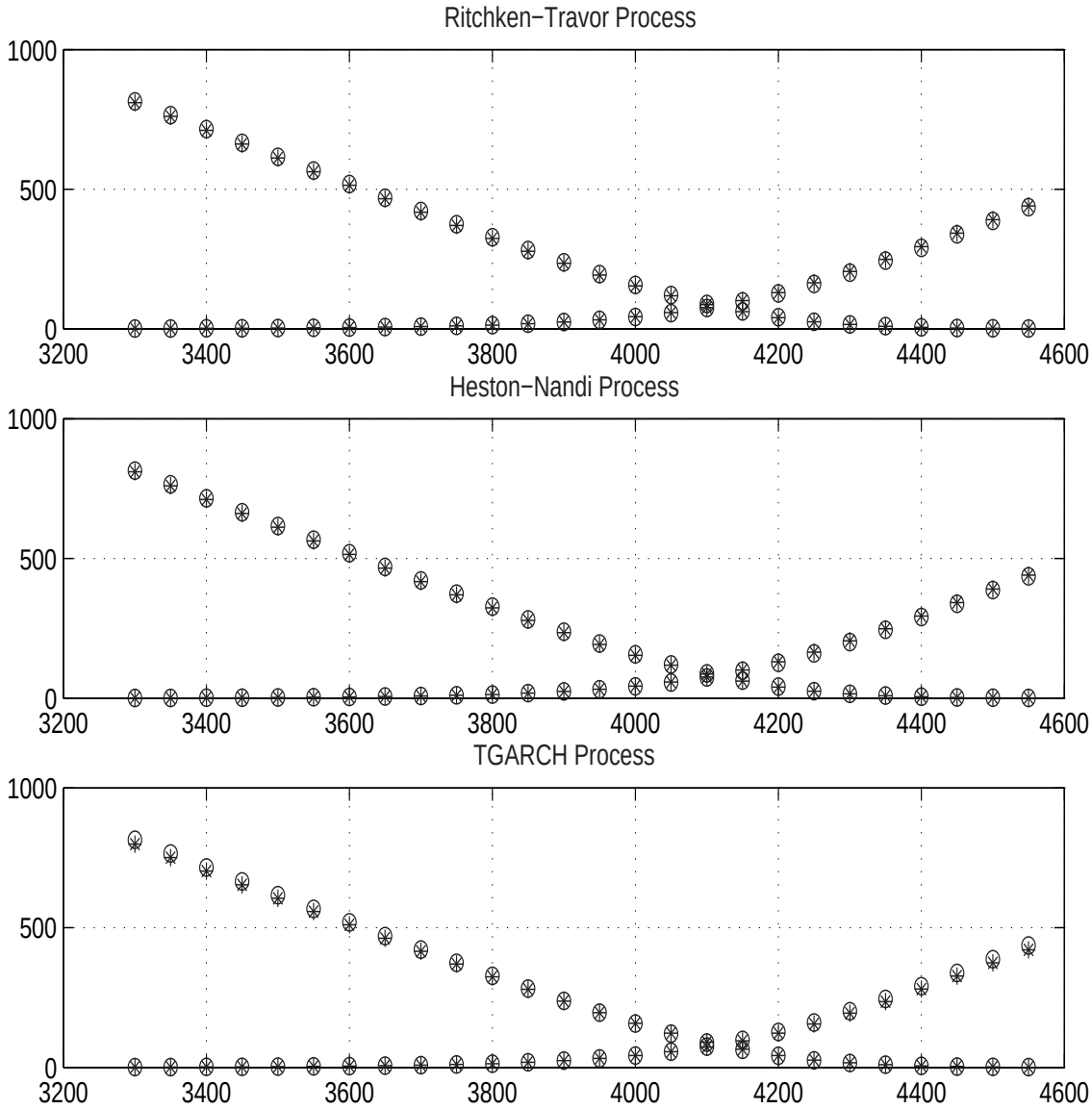

Figure 26: Empirical ($\circ$) and simulated ($*$) call and put option prices

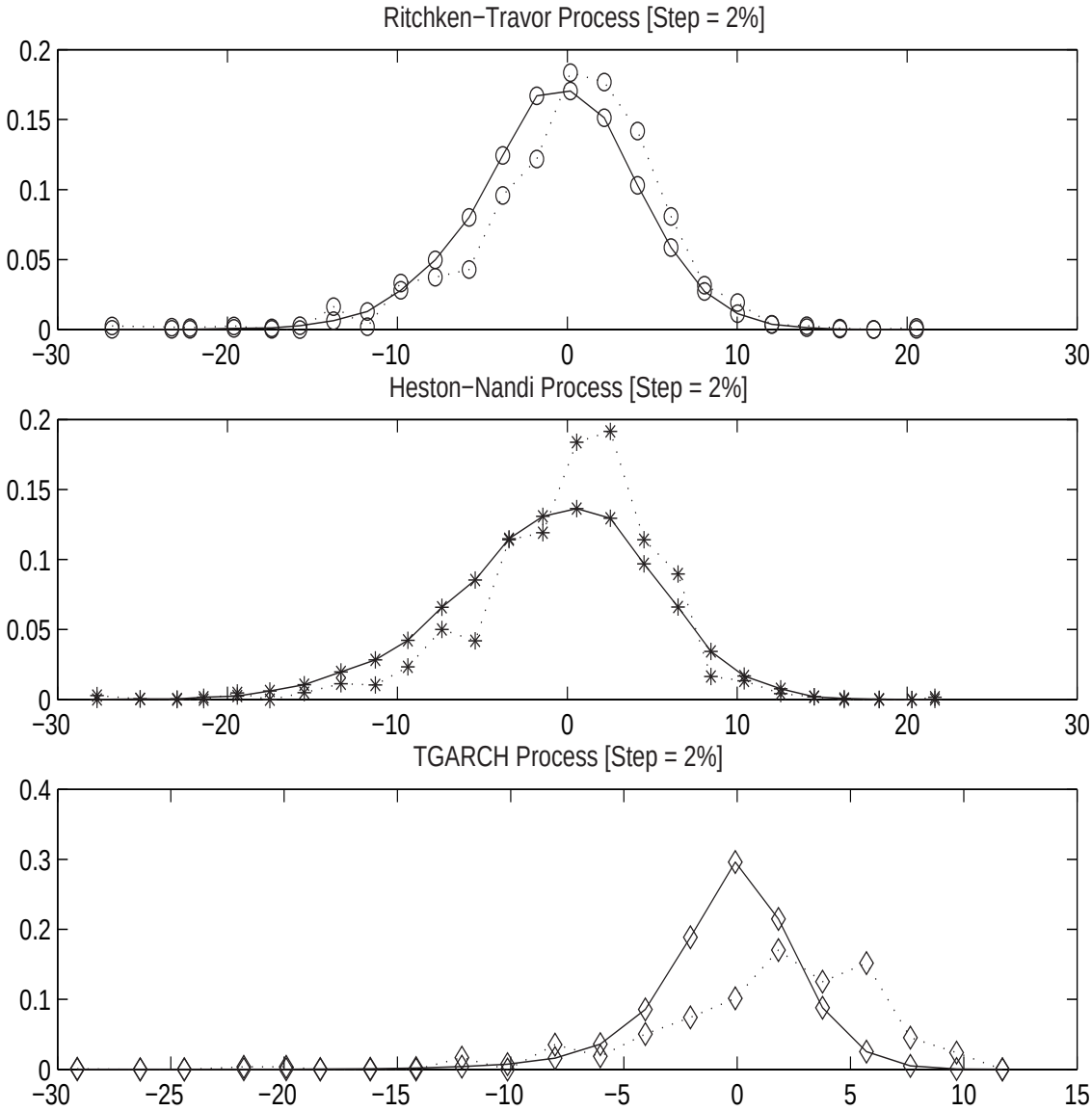

Figure 27: Objective (−) vs. risk-neutral (...) probability density of the underlying returns (using 26 call and 26 put on April 23, 2004)

8 Conclusions

I have proposed a general algorithm for the generation of underlying and contingent claim scenario trees, which can be used as input in a portfolio allocation model based on multi-stage stochastic programming.

The underlying scenario tree is computed by discretizing future price densities obtained from the simulation of GARCH processes. This allows us to avoid imposing any particular distribution structure for underlying returns and to consider volatility clustering, fat tails as well as leverage effects in our scenarios. Moreover, since the estimation of the GARCH models and the simulation of future price densities are based on the history of past prices, these are indirectly reflected in our scenarios.

The obtained underlying scenarios are then used to infer the risk neutral probabilities and option prices. Unlike past models present in literature, our algorithm shows a natural link between the objective and the risk-neutral world. We do not have to impose any utility function, nor a limited structure for the tree to infer risk neutral probabilities from the objective probabilities, or vice versa. In fact, the number of periods and nodes of our scenario trees is potentially unrestricted and can be completely determined simply by solving standard minimization problems. Moreover, the obtained option price scenarios do not reflect only information from a limited number of existing options and related implied volatilities, but also information from the underlying price process. We can therefore combine in a simple and consistent way the characteristics of underlying assets and the forward nature of existing market option prices.

The empirical results are quite encouraging. Although we consider the whole set of existing 52 call and put options on the DAX 100 index on April 23, 2004, the percentage errors and the average price errors are quite small. Hence, for deep out of the money options, our pricing model seems to work well. Moreover, using two different underlying prices GARCH process (The RT and the HN GARCH) and therefore for two different risk neutral-probability priors, the obtained option prices are exactly the same. We interpret this result as evidence for robustness of our model. Finally, if we compare the empirical four week price densities over the last three years of data with the simulated week four underlying scenario density using the RT GARCH process, we observe a relative closeness, which make us quite confident about the quality of the generated scenarios.

The main point of improvement for the algorithm is probably the simulation of the probability set. Because of the cited critiques about the Barone-Adesi et al. (1999) method, it would be interesting to introduce more sophisticated methods for the simulation of the underlying densities, like for example the one proposed by Audrino and Barone-Adesi (2005). Moreover, besides the approximation method used in our algorithm, there are many other scenario reduction and scenario tree construction algorithms all based on probability metrics, which are not analysed in this work (See for example Casey and Sen (2002), Kouwenberg (1986) or Gröwe-Kuska, Heitsch, and Römisch (2003) among others). A deeper analysis of such different models and the identification of their advantages and drawbacks for the introduction in an algorithm similar to the one introduced in this work could be of interest, but is left for future research.

References

Audrino, F., and G. Barone-Adesi, 2005, A Multivariate FGD Technique to Improve VaR Computation in Equity Markets, *Computational Management Science* 2, 87–106.

Audrino, F., and P. Bühlmann, 2003, Volatility Estimation with Functional Gradient Descent for Very High-Dimensional Financial Time Series, *Journal of Computational Finance* 6,3, 65–89.

Barone-Adesi, G., R. Engle, and L. Mancini, 2004, GARCH options in incomplete markets, Working Paper, Department of Finance, University of Lugano.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper, 1999, VaR without Correlations for Portfolios of Derivative Securities, *Journal of Futures Markets* 19,5, 583–602.

Birge, J.R., and F. Louveaux, 1997, *Introduction to Stochastic Programming.* (Springer New York).

Bollerslev, T., 1990, Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *The Review of Economics and Statistics* 72, 498–505.

Bradley, S.P., and D.B. Crane, 1972, A Dynamic Model for Bond Portfolio Management, *Management Science* 19,2, 139–151.

Breeden, D.T., and R.H. Litzenberger, 1978, Prices of State-Contingent Claims implicit in Option Prices, *Journal of Business* 51,4, 621–651.

Brodt, A.I., 1984, Intertemporal Bank Asset and Liability management, *Journal of Bank Research* 15, 82–94.

Carino, D.R., D.H. Myers, and W.T. Ziemba, 1998, Concepts, Technical Issues and Uses of the Russel-Yasuda Kasai Financial Planning Model, *Operation Research* 46, 450–462.

Carino, D.R., and W.T. Ziemba, 1998, Formulation of the Russel-Yasuda Kasai Financial Planning Model, *Operation Research* 46, 433–449.

Casey, M., and S. Sen, 2002, The Scenario Generation Algorithm for Multistage Stochastic Linear Programming, Dept. Systems and Industrial Engineering, University of Arizona.

Cohen, K.J., and F.S. Hammer, 1967, Planning for Liquidity in Financial Institutions: The Chance-Constrained Method, *Journal of Finance* 22, 42–61.

Cox, J., J. Ingersoll, and S. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385–407.

Cox, J., S. Ross, and M. Rubinstein, 1979, Option Pricing: A Simplified Approach, *Journal of Financial Economics* 7, 229–264.

Dempster, M.A.H., and A.M. Ireland, 1988, A Financial Expert Decision Support System, *In: Mitra G (ed.) Mathematical Models for Decision Support* NATO ASI Series Vol. F48, Springer, Berlin, 415–440.

Duan, J.C., 1995, The GARCH Option Pricing Model, *Mathematical Finance* 5,1, 13–32.

Dupačová, J., 1999, Portfolio Optimization via Stochastic Programming: Methods of Output Analysis, *Mathematical Methods of Operation Research* 50, 245–270.

Dupačová, J., N. Gröwe-Kuska, and W. Römisch, 2002, Scenario Reduction in Stochastic Programming, An Approach Using Probability Metrics, *Math. Programming, Ser.A* 95, 493–511.

Dupačová, J., and W. Römisch, 1998, Quantitative Stability for Scenario-Based Stochastic Programs. In:, *Prague Stochastics '98* (M. Hušková et al. eds.), JČMF Prague, 119–124.

Engle, R.F., and V.K. Ng, 1993, Measuring and Testing the Impact of News on Volatility, *Journal of Finance* 48,5, 1749–1778.

Feller, W., 1951, Two Singular Diffusion Processes, *Annals of Mathematics* 54, 173–182.

Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, On the Relation between the Expected Value and Volatility of the Nominal Excess Return on Stocks, *Journal of Finance* 48, 1779–1801.

Gröwe-Kuska, N., H. Heitsch, and W. Römisch, 2003, Scenario Reduction and Scenario Tree Construction for Power Management Problems, in IEEE Bologna Power Tech Proceedings.

Hester, D.D., and J.L. Pierce, 1975, *Bank Management and Portfolio Behavior*. (Yale University Press Conn.).

Heston, S.L., 1993, A Closed-Form Solution for Options with Stochastic Volatility, with Applications to Bond and Currency Options, *Review of Financial Studies* 6, 327–343.

Heston, S.L., and S. Nandi, 2000, A Closed-Form GARCH Option Valuation Model, *Review of Financial Studies* 13,3, 585–625.

Hochreiter, R., and G. Pflug, 2003, Scenario Generation for Multi-stage Decision models: an apporach based on multidimensional facility location, AURORA Technical Report, Department of Statistics and Decision Support Systems, University of Vienna.

Høyland, K., M. Kaut, and S.W. Wallace, 2003, A heuristic for moment-matching scenario generation, *Computational Optimization and Applications* 24, 169–185.

Høyland, K., and S.W. Wallace, 2001, Generating Scenario Trees for Multistage Decision Problems, *Management Science* 47,2, 295–307.

Jackwerth, J.C., and M. Rubinstein, 1996, Recovering Probability Distribution from Option Prices, *The Journal of Finance* 51,5, 1611–1631.

Keefer, D.L., 1994, Certainty Equivalents for Three-point Discrete-distribution Approximations, *Management Science* 40,6, 760–773.

Kouwenberg, R., 1986, Scenario Generation and Stochastic Programming Models for Asset Liability Management, *European Journal of Operational Research* 134, 51–64.

Kusy, M.I., and W.T. Ziemba, 1986, A Bank Asset and Liability Management Model, *Operation Research* 34,3, 356–376.

Laurent, A., 2003, *Derivatives and the Asset Allocation Decision: A Synthesis between Portfolio Diversification and Portfolio Insurance*. (PhD Dissertation, Difo-Druck GmbH Bamberg).

Miller, A.C., and T.R. Rice, 1983, Discrete Approximation of Probability Distribution, *Math. Programming, Ser.B* 30(4), 127–139.

Mulvey, J.M., and H. Vladimirou, 1989, Stochastic Network Optimization Models for Investment Planning, *Annals of Operation Research* 20, 187–217.

N., Gülpinar, B. Rustem, and R. Settergren, 2001, Simulation and Optimization Approaches to Scenario Tree Generation, Department of Computing, Imperial College of Science, Technology and Medicine.

Nelson, D.B., 1990, ARCH Models as Diffusion Approximations, *Journal of Econometrics* 45, 7–38.

Pflug, G.Ch., 2001, Scenario Tree Generation for Multiperiod Financial Optimization by Optimal Discretization, *Management Science* 29,3, 352–362.

Pritsker, M., 2001, The Hidden Dangers of Historical Simulation, Finance and Economics Discussion Series 2001-27, Washington: Board of Governors of the Federal Reserve System.

Rachev, S.T., 1991, *Probability Metrics and the Stability of Stochastic Models*. (J. Wiley and Sons New York).

Ritchken, P., and R. Travor, 1999, Pricing Options under Generalized GARCH and Stochastic Volatility Processes, *Journal of Finance* 54,1, 377–402.

Rubinstein, M., 1994, Implied Binomial Trees, *Journal of Finance* 69,3, 771–818.

Schyns, M., Y. Crama, and G. Hübner, 2003, Grafting Information in Scenario Trees Application to Option Prices, University of Liège, Working paper N° 200303.

Shimko, D., 1993, Bounds of Probability, *Risk* 6,4, 33–37.

Vallander, S.S., 1973, Calculation of the Wasserstein Distance between Probability Distributions on the Line, *Theor. Prob. Appl.* 18, 784–786.

Yu, L.Y., Y.D. Ji, and S.Y. Wang, 2003, Stochastic Programming Models in Financial Optimization: A Survey, *Advanced modeling and Optimization* 5,1, 1–26.